VOLUME 108 PART B NUMBER 41 SEPTEMBER 1961

The Proceedings

OF

THE INSTITUTION OF ELECTRICAL ENGINEERS

FOUNDED 1871: INCORPORATED BY ROYAL CHARTER 1921

PART B

ELECTRONIC AND COMMUNICATION ENGINEERING
(INCLUDING RADIO ENGINEERING)

Price Ten Shillings and Sixpence

THE INSTITUTION OF ELECTRICAL ENGINEERS

FOUNDED 1871 INCORPORATED BY ROYAL CHARTER 1921

PATRON: HER MAJESTY THE QUEEN

COUNCIL 1960–1961

President

SIR HAMISH D. MACLAREN, K.B.E., C.B., D.F.C.,* LL.D., B.Sc.

Past-Presidents

W. H. ECCLES, D.Sc., F.R.S.
THE RT. HON. THE EARL OF MOUNT EDGCUMBE, T.D.
J. M. DONALDSON, M.C.
PROF. E. W. MARCHANT, D.Sc.
H. T. YOUNG.
SIR GEORGE LEE, O.B.E., M.C.
J. R. BEARD, C.B.E., M.Sc.
SIR NOEL ASHBRIDGE, B.Sc.(Eng.).
SIR HARRY RAILING, D.Eng.
P. DUNSHEATH, C.B.E., M.A., D.Sc.(Eng.), LL.D.
SIR VINCENT Z. DE FERRANTI, M.C.
T. G. N. HALDANE, M.A.
PROF. E. B. MOULLIN, M.A., Sc.D., LL.D.
SIR ARCHIBALD J. GILL, B.Sc.(Eng.).
SIR JOHN HACKING.
COL. B. H. LEESON, C.B.E., T.D.
SIR HAROLD BISHOP, C.B.E., B.Sc.(Eng.), F.C.G.I.
SIR JOSIAH ECCLES, C.B.E., D.Sc.
THE RT. HON. THE LORD NELSON OF STAFFORD.
SIR W. GORDON RADLEY, K.C.B., C.B.E., Ph.D.(Eng.).
S. E. GOODALL, M.Sc.(Eng.), F.Q.M.C.
SIR WILLIS JACKSON, D.Sc., D.Eng., LL.D., F.R.S.

Vice-Presidents

B. DONKIN, B.A. O. W. HUMPHREYS, C.B.E., B.Sc. G. S. C. LUCAS, O.B.E., F.C.G.I. C. T. MELLING, C.B.E., M.Sc.Tech. A. H. MUMFORD, O.B.E., B.Sc.(Eng.).

Honorary Treasurer

C. E. STRONG, O.B.E., B.A., B.A.I.

Ordinary Members of the Council

J. C. ARKLESS, B.Sc.
PROF. H. E. M. BARLOW, Ph.D., B.Sc.(Eng.), F.R.S.
D. A. BARRON, M.Sc.
C. O. BOYSE, B.Sc.(Eng.).
F. H. S. BROWN, C.B.E., B.Sc.
PROF. M. W. HUMPHREY DAVIES, M.Sc.
SIR JOHN DEAN, B.Sc.
L. DRUCQUER.
J. M. FERGUSON, B.Sc.(Eng.).
D. C. FLACK, B.Sc.(Eng.), Ph.D.
R. J. HALSEY, C.M.G., B.Sc.(Eng.), F.C.G.I.
R. A. HORE, M.A., B.Sc., Ph.D.
J. S. MCCULLOCH.
PROF. J. M. MEEK, D.Eng.
THE HON. H. G. NELSON, M.A.
H. V. PUGH.
J. R. RYLANDS, M.Sc., J.P.
R. L. SMITH-ROSE, C.B.E., D.Sc., Ph.D
G. A. V. SOWTER, Ph.D., B.Sc.(Eng.).
H. G. TAYLOR, D.Sc.(Eng.).
D. H. TOMPSETT, B.Sc.(Eng.).

Chairmen and Past-Chairmen of Sections

Electronics and Communications:
T. B. D. TERRONI, B.Sc.
†M. J. L. PULLING, C.B.E., M.A.

Measurement and Control:
C. G. GARTON.
†PROF. A. TUSTIN, M.Sc.

Supply:
J. E. L. ROBINSON, M.Sc.
†J. R. MORTLOCK, Ph.D., B.Sc.(Eng.).

Utilization:
J. M. FERGUSON, B.Sc.(Eng.).
†T. E. HOUGHTON, M.Eng.

Chairmen and Past-Chairmen of Local Centres

East Midland Centre:
LT.-COL. W. E. GILL, T.D.
†D. H. PARRY, B.Sc.

North-Western Centre:
F. LINLEY.
†F. J. HUTCHINSON, M.Eng.

Mersey and North Wales Centre:
D. A. PICKEN.
†T. A. P. COLLEDGE, B.Sc.(Eng.).

Northern Ireland Centre:
J. MCA. IRONS.
†T. S. WYLIE.

Southern Centre:
R. GOFORD.
†W. D. MALLINSON, B.Sc.(Eng.).

North-Eastern Centre:
D. H. THOMAS, M.Sc.Tech., B.Sc.(Eng.).
†H. WATSON-JONES, M.Eng.

Scottish Centre:
R. B. ANDERSON.
†J. A. AKED, M.B.E.

Western Centre:
A. C. THIRTLE.
†H. JACKSON, B.Sc.(Eng.).

North Midland Centre:
F. W. FLETCHER.
†PROF. G. W. CARTER, M.A.

South Midland Centre:
BRIGADIER F. JONES, C.B.E., M.Sc.
†G. F. PEIRSON.

† Past Chairman.

ELECTRONICS AND COMMUNICATIONS SECTION COMMITTEE 1960–1961

Chairman

T. B. D. TERRONI, B.Sc.

Vice-Chairmen

R. J. HALSEY, C.M.G., B.Sc.(Eng.), F.C.G.I. J. A. RATCLIFFE, C.B.E., M.A., F.R.S.

Past-Chairmen

G. MILLINGTON, M.A., B.Sc. M. J. L. PULLING, C.B.E., M.A.

Ordinary Members of Committee

W. H. ALDOUS, B.Sc., D.I.C.
D. A. BARRON, M.Sc.
P. A. T. BEVAN, C.B.E., B.Sc.
J. BROWN, M.A., Ph.D.
G. G. GOURIET.
J. MOIR.
L. J. I. NICKELS, B.Sc.(Eng.).
W. J. PERKINS.
N. C. ROLFE, B.Sc.(Eng.).
K. F. SANDER, M.A., Ph.D., B.Sc.
J. A. SAXTON, D.Sc., Ph.D.
T. R. SCOTT, D.F.C., B.Sc.
F. J. D. TAYLOR, O.B.E., B.Sc.(Eng.).
R. C. G. WILLIAMS, Ph.D., B.Sc.(Eng.).
R. C. WINTON, B.Sc.
A. J. YOUNG, B.Sc.(Eng.).

And

The President (*ex officio*).
The Chairman of the Papers Committee.
PROFESSOR H. E. M. BARLOW, Ph.D., B.Sc.(Eng.), F.R.S. (representing the Council).
E. H. COOKE-YARBOROUGH, M.A. (Co-opted Member).
E. D. TAYLOR, M.Sc. (representing the North-Eastern Measurement and Electronics Group).
H. V. BECK, B.Sc., M.A. (representing the Cambridge Electronics and Measurement Group).
S. D. MELLOR, B.Eng. (representing the North-Western Electronics and Communications Group).
J. R. POLLARD, M.A. (representing the East Midland Electronics and Control Group).
J. STEWART, M.A., B.Sc. (representing the Scottish Electronics and Measurement Group).
R. E. YOUNG, B.Sc.(Eng.) (representing the South Midland Electronics and Measurement Group).

The following nominees:
Royal Navy: CAPTAIN J. S. RAVEN, B.Sc., R.N.
Army: COL. R. G. MILLER, M.A.
Royal Air Force: GROUP CAPTAIN D. W. ROWSON, B.Sc.(Eng.), R.A.F.

MEASUREMENT AND CONTROL SECTION COMMITTEE 1960–1961

Chairman

C. G. GARTON.

Vice-Chairmen

W. S. ELLIOTT, M.A.; A. J. MADDOCK, D.Sc.

Past-Chairmen

PROFESSOR A. TUSTIN, M.Sc.; J. K. WEBB, M.Sc.(Eng.), B.Sc.Tech.

Ordinary Members of Committee

S. S. CARLISLE, M.Sc.
W. J. JEFFERSON.
C. A. LAWS.
A. C. LYNCH, M.A., B.Sc.
R. E. MARTIN.
A. NEMET, Dr.Sc.Techn.
S. N. POCOCK.
W. RENWICK, M.A., B.Sc.
R. D. TROTTER, B.Sc.(Eng.).
J. H. WESTCOTT, B.Sc.(Eng.), Ph.
F. C. WIDDIS, B.Sc.(Eng.), Ph.D.D.

And

The President (*ex officio*).
The Chairman of the Papers Committee.
G. A. V. SOWTER, Ph.D., B.Sc.(Eng.) (representing the Council).
C. C. BAXENDALE (representing the North-Eastern Measurement and Electronics Group).
A. CHORLTON, B.Sc.Tech. (representing the North-Western Measurement and Control Group).
H. M. GALE, B.Sc.(Eng.) (representing the South Midland Electronics and Measurement Group).
D. R. HARDY, Ph.D., M.Sc.(Eng.) (representing the East Midland Electronics and Control Group).
W. H. P. LESLIE, B.Sc. (representing the Scottish Electronics and Measurement Group).
D. L. A. BARBER, B.Sc.(Eng.) (nominated by the National Physical Laboratory).

Secretary

W. K. BRASHER, C.B.E., M.A., M.I.E.E.

Deputy Secretary

F. JERVIS SMITH, M.I.E.E.

Principal Assistant Secretary

F. C. HARRIS.

Editor-in-Chief

G. E. WILLIAMS, B.Sc.(Eng.), M.I.E.E.

Before specifying magnetic materials consult this record

HIGHEST μ OBTAINABLE
. . . is in nickel-iron alloys—available in all forms down to ultra-thin strip.

SQUARE HYSTERESIS LOOP
. . . nickel-iron alloys are the best materials for magnetic amplifiers and saturable reactors.

LOW CURIE POINT
associated with certain nickel alloys provides a temperature dependant permeability—a valuable characteristic for compensating and control devices.

HIGH MAGNETOSTRICTION
Nickel and nickel alloys make the most rugged and efficient transducers for ultrasonic equipment.

HIGH B H. MAX
Nickel-cobalt-aluminium-iron permanent magnets provide the maximum energy per unit volume, extreme stability and the greatest resistance to the effects of temperature change and vibration.

Design with Nickel-containing MAGNETIC MATERIALS

Send for a free publication 'Nickel-containing Magnetic Materials'

THE INTERNATIONAL NICKEL COMPANY (MOND) LIMITED THAMES HOUSE MILLBANK LONDON SW1

TGA GN50

NEW CINTEL ALL & SELF

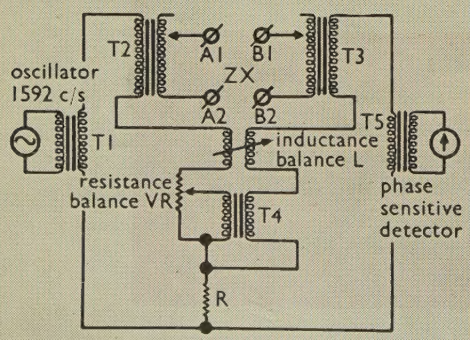
Accurate within ±one per cent of full scale on the inductance and resistance ranges ·0·005μH to 30mH, and ·500μΩ to 3300Ω respectively.

THE BRIDGE is an all transistor version of the well-known valve model of the same name, which it now supersedes.

This new version retains the same rigid specification with the added advantages of greater scale length, increased accuracy, improved reliability, printed circuit construction, simpler and quicker operation from either mains or battery, and no warm-up period.

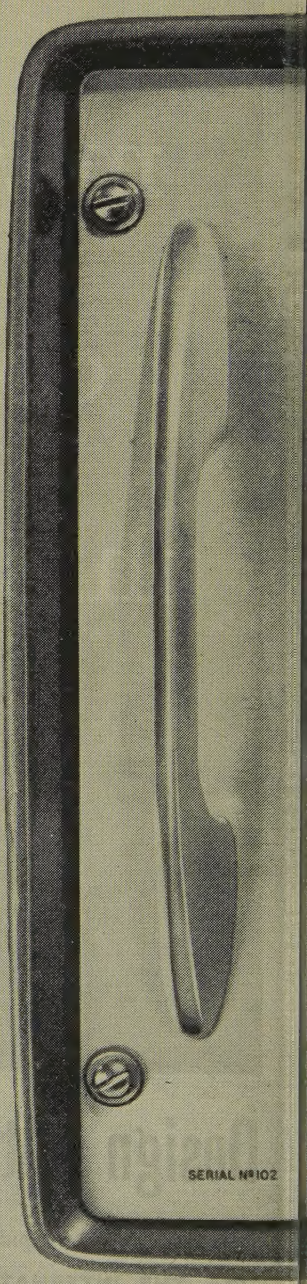
APPLICATION. The instrument has been designed essentially to meet the demand for a bridge capable of accurate low self or mutual inductance with a wide variety of applications.

It may be used to measure the a.f. Inductance of small coils which have a low Q value at audio frequencies, such as radio and television receiver tuning coils. It can also be applied to the measurement of correcting inductors, loading coils, the inductance of transmission lines, cables and circuit 'strays'. This latter application is particularly useful in high frequency circuits. The bridge may also be used for measuring any mutual inductance within its range; for example, that of I.F. transformers. A particularly useful application of mutual measurement is in the setting up of artificial transmission lines. Measurements can be made step by step along the completed line, each section being treated as a three-terminal network. Mutual resistance can also be measured with precision.



A further use of the bridge is for the accurate measurement of low resistance values, the first scale reading being 100 microhms.

A complete technical description is available on request to . . .



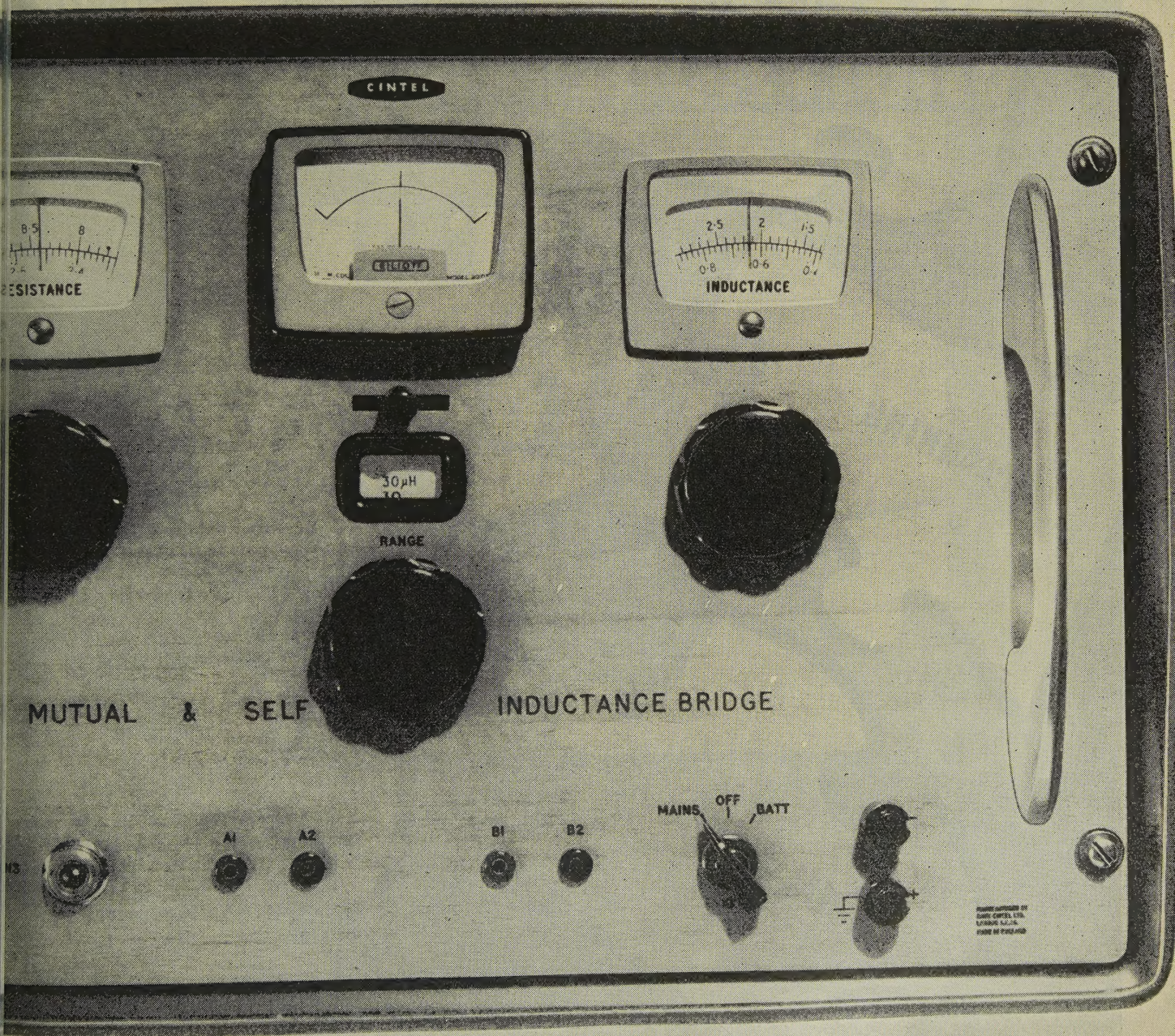
CINTEL instruments division

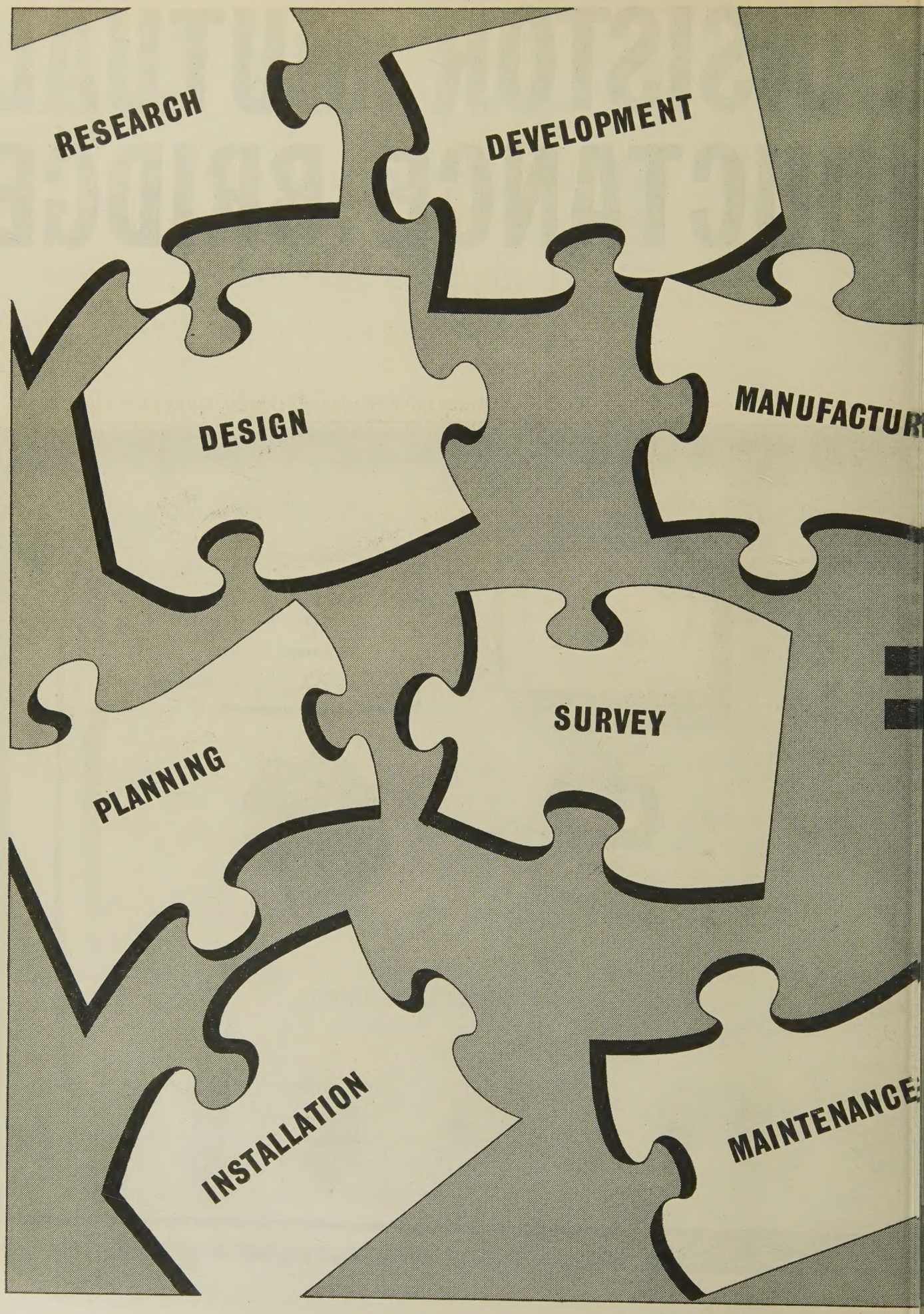
RANK CINTEL LIMITED

Worsley Bridge Road, Lower Sydenham,
LONDON, S.E.26. HITHER GREEN 4600

RANSISTOR MUTUAL
NDUCTANCE BRIDGE

Dimensions: 20·5″ wide x 14″ high x 13″ deep (52 x 35 x 33 cm).




RESEARCH
DEVELOPMENT
DESIGN
MANUFACTUR
SURVEY
PLANNING
INSTALLATION
MAINTENANCE

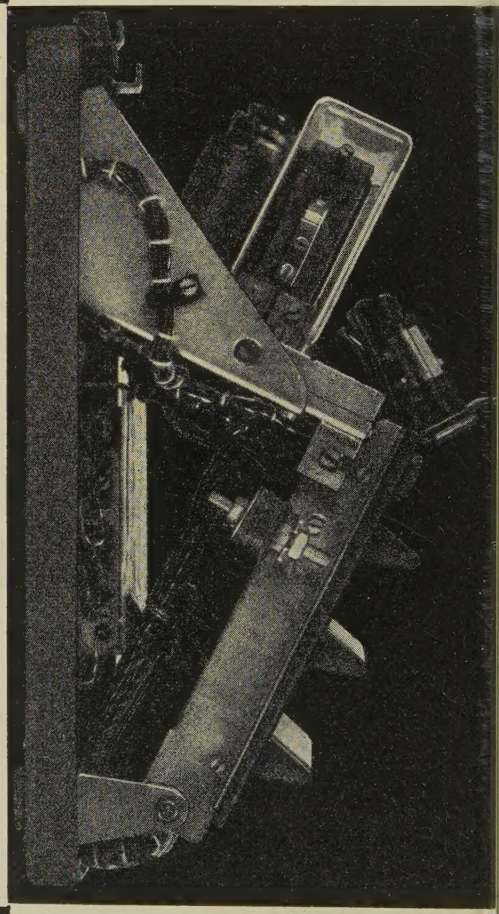
consult

TRANSMISSION DIVISION

THE GENERAL ELECTRIC COMPANY LIMITED OF ENGLAND

Telephone Works · Coventry · England *Works at Coventry · Middlesbrough · London · Portsmouth*

NEW STYLE CORDLESS SWITCHBOARD



Attractive appearance is an essential consideration when designing subscribers' apparatus, but important, too, is small size. Both these objectives have been achieved in this *lamp signalling* switchboard, designed in conjunction with the B.P.O. to supersede the existing indicator signalling 2 + 4 switchboard housed in a bulky, outmoded wooden cabinet.

Although considerably smaller than its predecessor, it provides two more extensions and has a total capacity for two exchange lines and six extensions.

Much of the reduction in size has been obtained by the adoption of a new 4-wire principle for local extension lines. This has enabled certain additional facilities to be provided with fewer components, as for example, operator re-call and secretarial 'hold'.

A grey plastic drop-on cover, with a simple release action, provides easy access to all components conveniently arranged on a 3-section, hinged chassis of drawn steel. The key and lamp panel is tastefully finished in a durable coating of grey P.V.C. and the ivory coloured key handles are shaped and marked to facilitate operation.

Power for transmission and signalling is normally derived from a mains driven unit.

ERICSSON TELEPHONES LTD
ETELCO LTD

Head Office :—22 Lincoln's Inn Fields, London, W.C.2. Tel. HOLborn 693

Crossed-Field Microwave Devices

Editor in Chief: E. OKRESS, *Sperry Gyroscope Company, Great Neck, New York*

Editors: G. MOURIER, *Compagnie Générale de Télégraphie Sans Fil, Orsay, France*

J. FEINSTEIN, *S-F-D Laboratories, Inc., Union, New Jersey*

E. KETTLEWELL, *General Electric Company Ltd., Wembley, England*

Assistant Editor for Volume 1: G. R. FEASTER, *Westinghouse Electric Corporation, Elmira, New York*

These two volumes fill an important gap in the literature by summarizing the progress and state of the art of crossed-field microwave devices in theory and practice. In order to indicate not only established uses but also possible future applications, the authors have included thorough surveys of the available literature. Their primary objective has been to enable students and professional physicists and engineers to understand the fundamentals of all existing important types of crossed-field devices, to perceive the problems posed by each element of the devices in practice and in theory, to know the essentials of the means available for solving these problems, and to develop approaches to the design of other devices. There is no other work which covers the subject so completely.

VOLUME 1, Principal Elements of Crossed-Field Devices

June 1961, 648 pp., 330 illustrations, 157s.

Introduction
By G. MOURIER *and* E. OKRESS

Periodic Structures
By J. ARNAUD, R. R. MOATS, M. C. PEASE, *and* MASAO NISHIMAKI

The Cathode Gun and Its Static Characteristics
By G. R. FEASTER, G. B. GAINES, D. L. GOLDWATER, W. R. HAYTER, JR., J. T. LAW, C. P. LEA-WILSON, A. J. MONK, M. C. PEASE, O. DOEHLER, *and* G. S. KINO

Dynamic Phenomena: Noise and Space-Charge Modes
By G. D. SIMS, O. BUNEMAN, GUNNAR HOK, R. L. JEPSEN, J. A. BRADSHAW, JOHN M. OSEPCHUK, S. OKAMURA, T. VAN DUZER, *and* J. R. WHINNERY

Dynamic Phenomena: Interaction of Beams and Circuits
By O. BUNEMAN, G. MOURIER, OM P. GANDHI, JOSEPH E. ROWE, JOSEPH F. HULL, J. FEINSTEIN, G. S. KINO, JAMES W. SEDIN, G. NOVICK, T. SHIMIZU, MASAO NISHIMAKI, *and* ROBERT DUNSMUIR

APPENDIX TO VOLUME 1

SUBJECT INDEX.

VOLUME 2, Principal Types of Crossed-Field Devices Analysis of Oscillator System Performance Regional Progress and Trends

June 1961, 520 pp., 370 illustrations, 128s. 6d.

PRINCIPAL TYPES OF CROSSED-FIELD DEVICES

Injection Type Tubes
By O. DOEHLER

Voltage Tuned Oscillators
By O. DOEHLER, D. A. WILBUR, P. H. PETERS, JR., C. LOUIS CUCCIA, *and* B. AGDUR

Mechanically Tuned Oscillators
By J. FEINSTEIN, R. J. COLLIER, F. E. VACCARO, M. J. BERNSTEIN, *and* N. M. KROLL

Amplifiers: Wide Band and Externally Stabilized Tunable Oscillators
By O. DOEHLER, W. C. BROWN, J. FEINSTEIN, R. J. COLLIER, *and* D. CHEN

Fixed Frequency Magnetron Oscillators
By M. J. BERNSTEIN, N. M. KROLL, H. A. H. BOOT, A. H. PICKERING, J. F. HULL, J. R. M. VAUGHAN, R. G. ROBERTSHAW, W. E. WILLSHAW, GORDON E. BECKER, *and* E. KETTLEWELL

ANALYSIS OF OSCILLATOR SYSTEM PERFORMANCE

Phasing by RF Signals
By EDWARD E. DAVID, JR.

Frequency Pushing
By C. R. SCHUMACHER

Loading Effects
By WILBUR L. PRITCHARD

Frequency Modulation
By J. S. DONAL, JR.

Amplitude Modulation
By J. S. DONAL, JR.

Spectrum Shape
By C. R. SCHUMACHER

Starting Phenomena and Jitter
By G. C. TURRELL

REGIONAL PROGRESS AND TRENDS

Status in the U.S.A.
By J. FEINSTEIN *and* H. W. WELCH, JR.

Status in France
By P. GUENARD

Status in Japan
By K. MORITA

SUBJECT INDEX.

ACADEMIC PRESS, New York and London

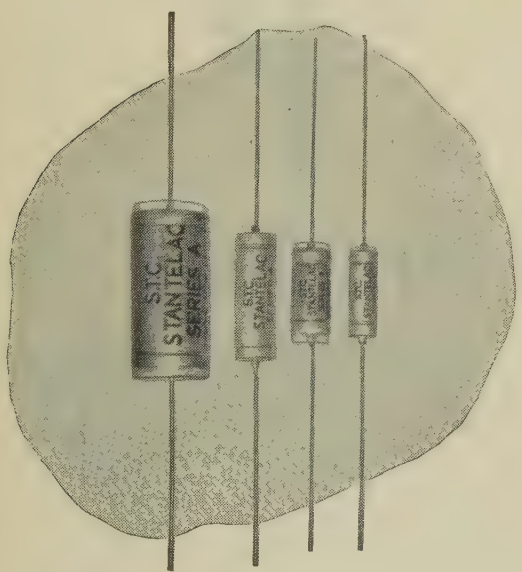
111 Fifth Avenue, New York 3 **17 Old Queen Street, London, S.W.1**

STC

'STANTELAC'

COATED FOIL CAPACITORS

a new development in Capacitor manufacturing techniques



EXCELLENT ELECTRICAL CHARACTERISTICS

HIGH STABILITY PERFORMANCE

SMALL PHYSICAL SIZE

LOW WORKING VOLTAGES
FOR TRANSISTOR CIRCUITS

TEMPERATURE RANGE —40°C TO +70°C.
H.I. CLIMATIC PROTECTION

CAPACITANCE RANGE 0·01μF TO 2·2μF

Send for details of the Stantelac range of capacitors and life test performance —

STC COMPONENTS GROUP

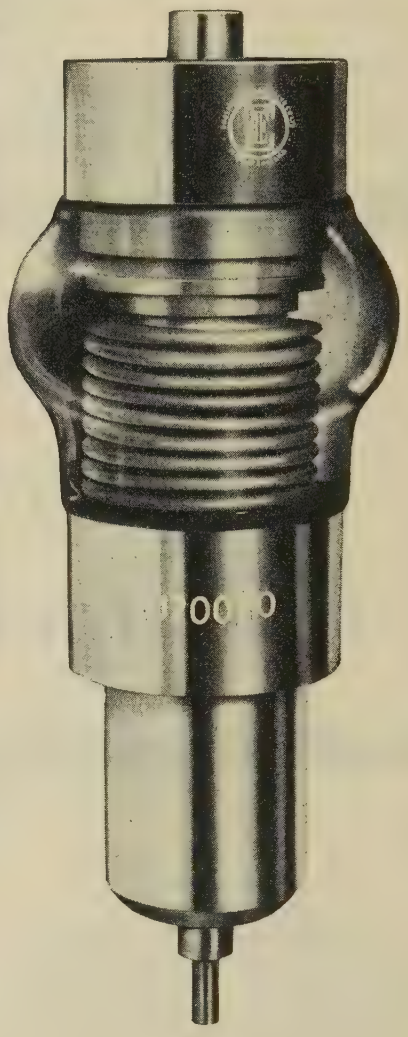
61/4MC

Standard Telephones and Cables Limited

CAPACITOR DIVISION: BRIXHAM ROAD · PAIGNTON · DEVON

'ENGLISH ELECTRIC' announce the first

HIGH VACUUM VARIABLE CAPACITORS



developed and manufactured in Britain

The range comprises five types for operation in high voltage r.f. circuits and all are tunable over an approximately linear capacitance range. High vacuum variable capacitors offer outstanding advantages over conventional air dielectric counterparts:—

- *** Compactness relative to high capacitance and operating voltage.**
- *** Low self inductance and stray capacitance.**
- *** No electrostatic dust precipitation on plates.**
- *** Easily demountable.**

Full information on the present range is available from the address below:

Further types will be added to meet future requirements.

'ENGLISH ELECTRIC'

E.E.V. type	Approx linear capacitance range (pF)	Shaft turns in linear capacitance range	Max peak r.f. voltage (kV)	Max r.f. current (r.m.s.) (A)	Max length (in)	Max dia. (in)
U30/15	5—30	10.4	15	10*	6.5	2.13
U50/15	8—50	10.4	15	15*	6.5	2.75
U80/15	16—80	10.4	15	20*	6.5	3.30
U200/8	5.5—206	17	8	20†	9.5	2.49
U200/10	5.5—206	17	10	20†	9.5	3.50

* Up to 30 Mc/s † Up to 20 Mc/s

ENGLISH ELECTRIC VALVE COMPANY LIMITED

AGENTS THROUGHOUT THE WORLD | *Chelmsford, England. Telephone: Chelmsford 3491*

AP172

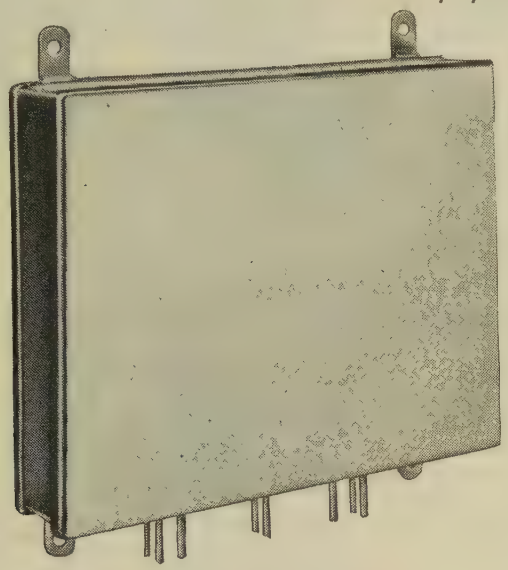
STC

offer packaged selectivity with

NEW MINIATURE

100 kc/s SINGLE SIDE-BAND CRYSTAL FILTER

for independent S.S.B. communications equipment

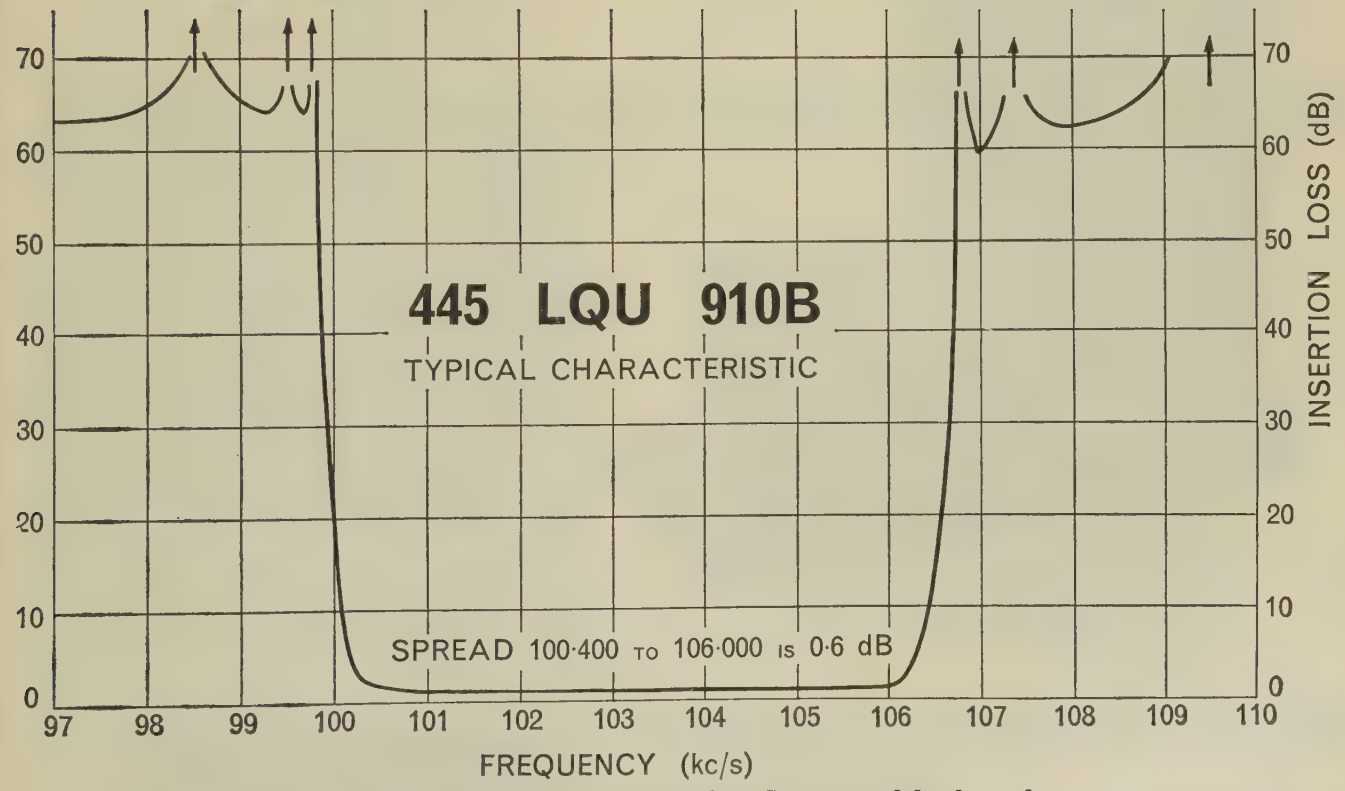


★ 3·95in. x 3·5in. x 1·02in.
(10cm x 8,9cm x 2,7cm)

★ Good carrier suppression

★ Small insertion loss

★ Minimum side-band ripple



Filters also available for lower side-band

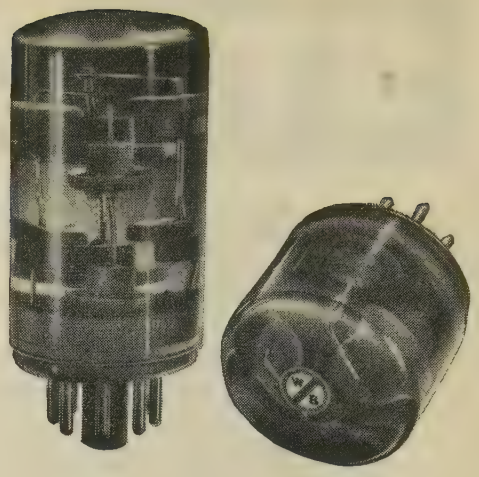
COMPONENTS GROUP

61/5MQ

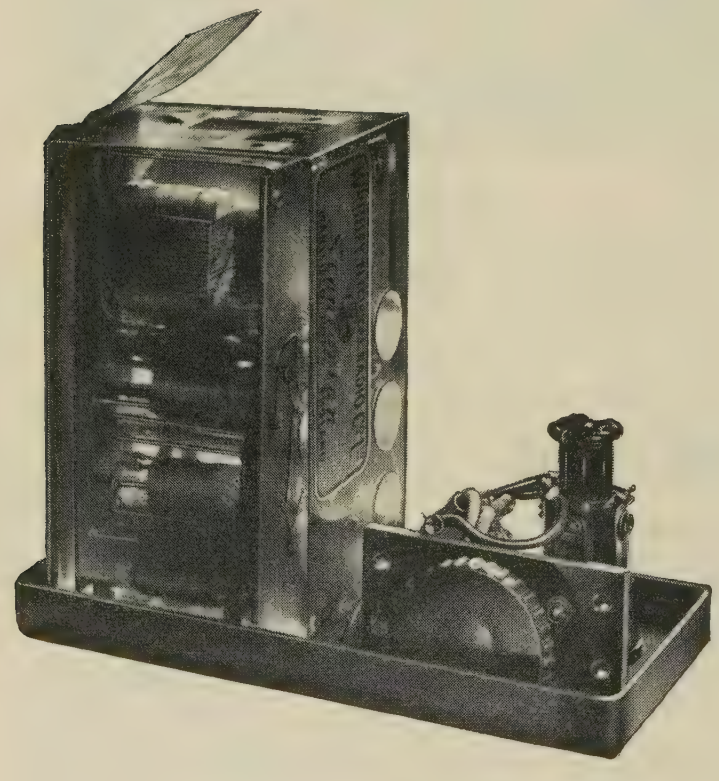
Standard Telephones and Cables Limited

QUARTZ CRYSTAL DIVISION: HARLOW · ESSEX

Comprehensive Cover



In literature describing their fault-locating equipment for Post Office use, Whiteley Electrical Radio Co. Ltd. have this to say about Araldite:



"The Araldite epoxy resins which we use provide mechanical and climatic protection, combined with excellent electrical insulation, for equipment used in extremes of temperature and humidity. Low shrinkage, excellent mechanical and electrical properties, and strong adhesion to components and leads afford positive protection at all times.

Such is the versatility of these resins that metals, ceramics, glass, mica, and laminated plastics can all be successfully bonded, giving exceptional strength and durability. A further important feature is the reduction in size and weight that can be achieved by encapsulation; while the characteristics of the resins provide complete waterproofing".

May we send you information regarding other uses of these remarkable epoxy resins?

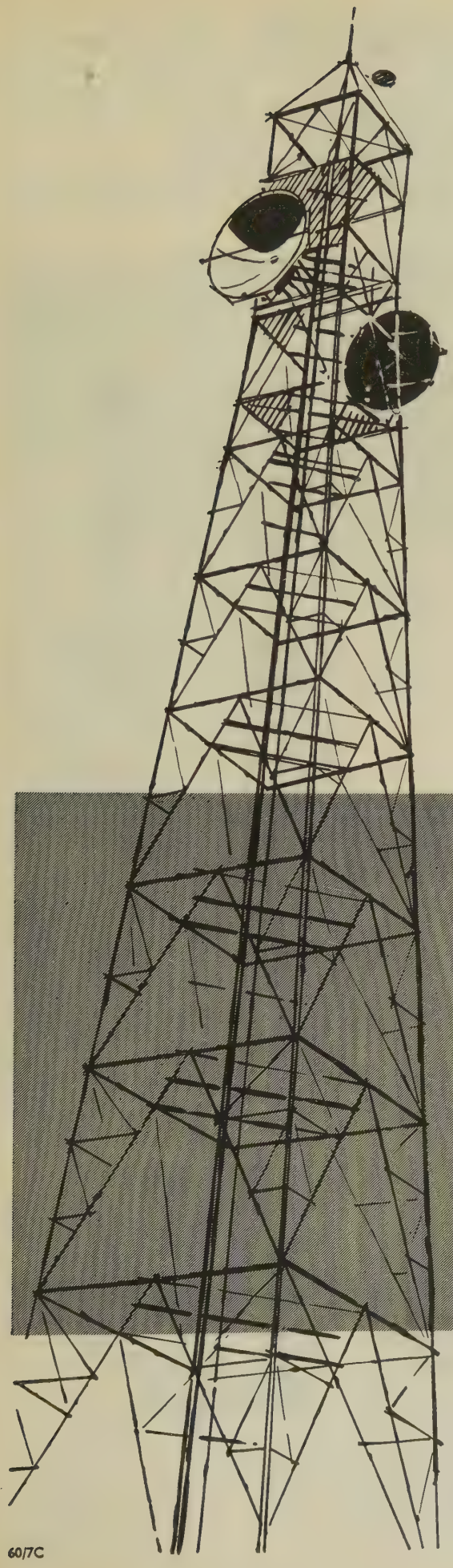
Araldite *epoxy resins*

Araldite is a registered trade mark

CIBA (A.R.L.) LIMITED

Duxford, Cambridge. Telephone: Sawston 2121

AP531



STC

PLAN, DESIGN, MANUFACTURE AND INSTALL COMPLETE S.H.F. SYSTEMS FOR TELEPHONE AND TELEVISION TRANSMISSION

STC can supply microwave radio systems complete with the necessary cable extensions and frequency division multiplex equipment.

STC EXPERIENCE

STC have supplied 4000 Mc/s systems having a capacity of over 4½ MILLION Telephone circuit miles and over 5000 Television channel miles.

STC are supplying systems to 18 countries

4000 Mc/s Terminal equipment cubicles.

STC COMMUNICATIONS GROUP

Experienced STC engineers are ready to help solve your communications problems.

Standard Telephones and Cables Limited

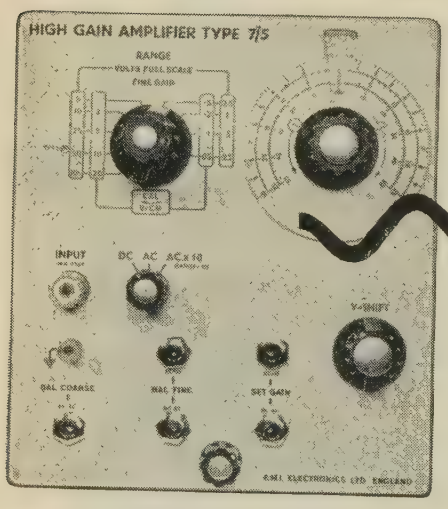
TRANSMISSION DIVISION: NORTH WOOLWICH · LONDON · E.16

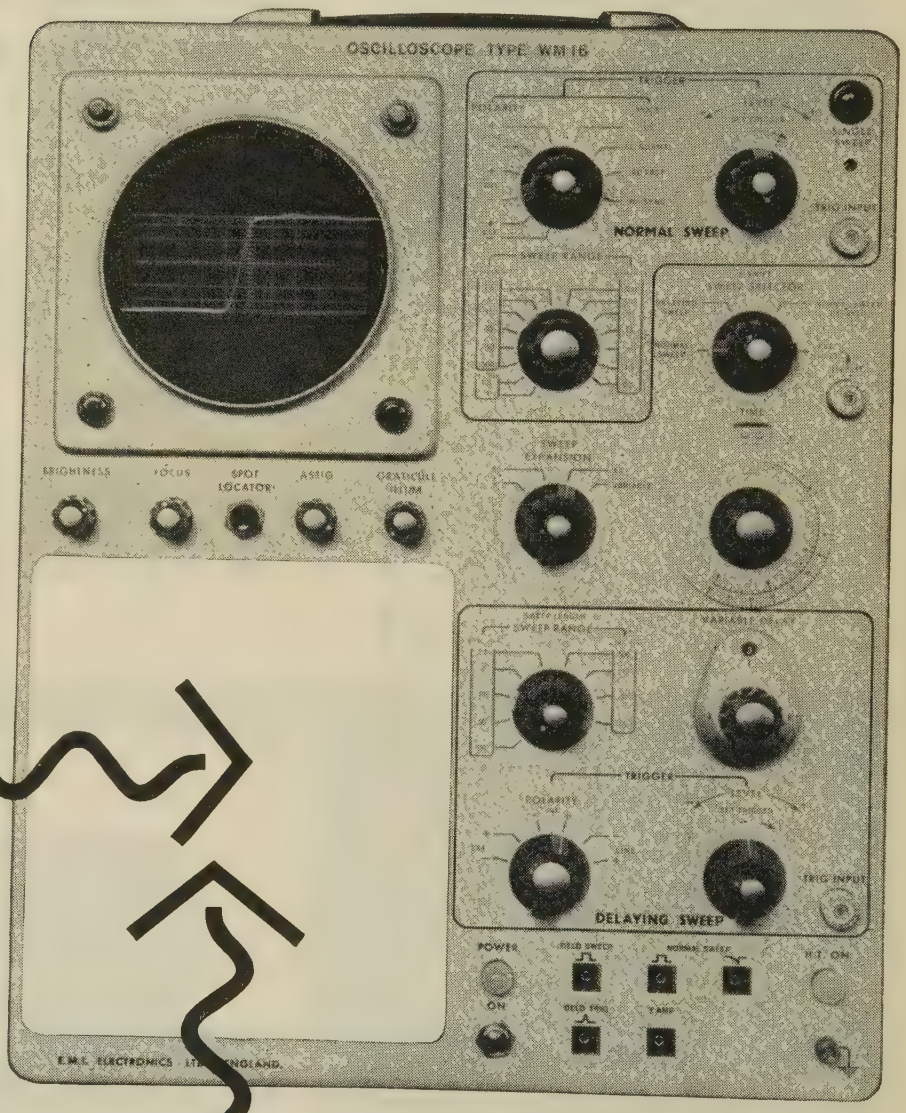
60/7C

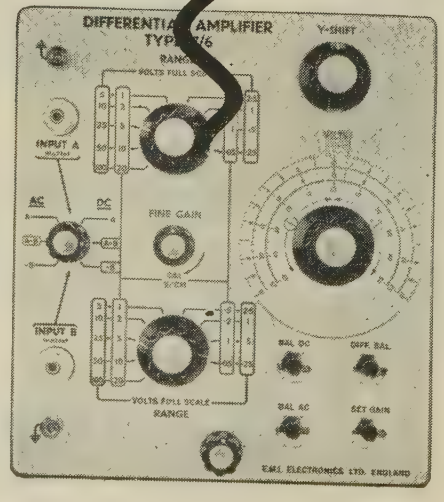
Still more versatility for the WM16! Even before this, no other oscilloscope in the same price range could equal its performance. Proved in action for over a year in government establishments, universities and industrial laboratories, the WM16 has shown itself ideal for radar, television, computers and millimicro-second oscillography, as well as for general laboratory electronic work.

Now the WM16 is given even greater versatility by the addition of 2 new plug-in units, which establish it even more firmly in a class of its own.

2 NEW UNITS FOR THIS HIGH PERFORMANCE 'SCOPE







High Gain Amplifier Type 7/5 *(above)*
High Gain, 5 mV/cm, 5 c/s – 25 Mc/s
Normal Gain, 50 mV/cm, DC – 40 Mc/s

Differential Amplifier Type 7/6 *(right)*
Two inputs can be displayed either separately or differentially.
Bandwidth DC – 25 Mc/s
Max. sensitivity 50 mV/cm
Rejection ratio greater than 100 : 1

General features of the WM16

Measurement accuracy	3%
Sweep delay	1 μsec – 150 m sec
Normal Sweep rate	12·5 m μsec/cm – 0·5 sec/cm

EMI

Ask now for technical information or a demonstration of the WM16 and its new plug-in units.

EMI ELECTRONICS LTD

INSTRUMENT DIVISION, HAYES, MIDDX
TELEPHONE: HAYES 3888 EXT. 2223

EE159

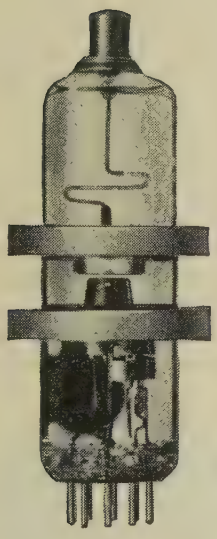
STC MICROWAVE OSCILLATORS

Coaxial Line Oscillators V types — Excellent frequency stability. High degree of modulation linearity.

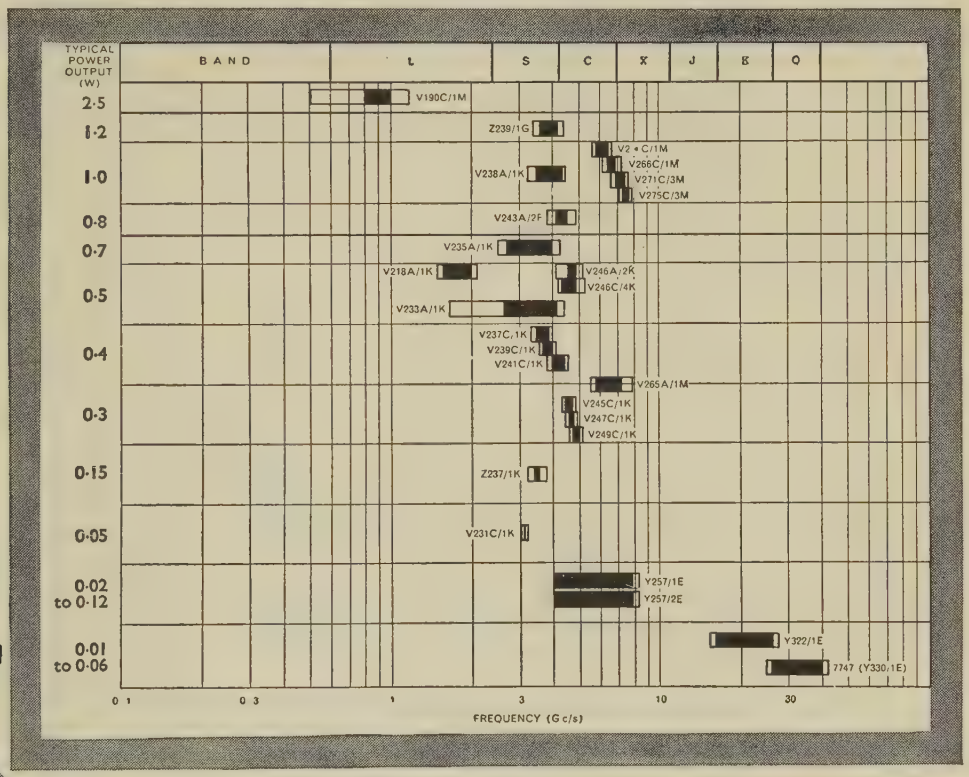
H-wave Oscillators V types — Low working voltage. No forced air cooling.

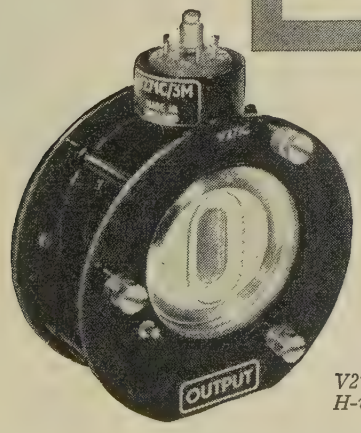
Backward Wave Oscillators Y types — Very wide electronic frequency coverage.

Reflex Klystrons Z types — Wide electronic tuning range. High degree of modulation linearity.



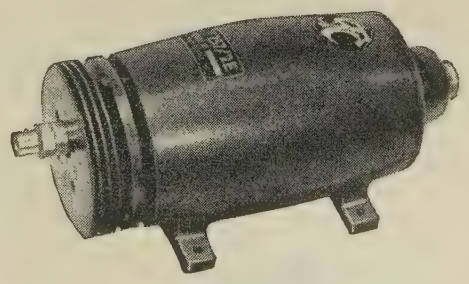
Z237/1K





V271C/3M
H-wave Oscillator

Send for a copy of the new edition of the illustrated brochure "STC Microwave Tubes" MS/113.



Y257/2E

COMMUNICATIONS GROUP

61/6MS

Standard Telephones and Cables Limited

VALVE DIVISION: FOOTSCRAY • SIDCUP • KENT

Cossorscope model 1076

SPECIFICATION

CATHODE-RAY TUBE: 5 in. (12.7 cm) diameter screen, single-beam, operating at 10 kV. Green fluorescence—phosphor P1. Phosphors P7 and P11 available to special order. Display area 10 cm (horizontal) by 6 cm (vertical).

Y AMPLIFIER: Basic Model 1076; d.c. to 80 Mc/s (30% down); rise-time 5 mμsec; sensitivity 200 mV/cm. With plug-in unit Model 1078; rise-time 6 mμsec to 7 mμsec; 50 mV/cm to 50 V/cm calibrated. With plug-in unit Model 1080; d.c. to 1 Mc/s (30% down); 1 mV/cm to 50 V/cm calibrated (differential inputs). With plug-in unit Model 1081; 3 c/s to 40 Mc/s (30% down); rise-time 10 mμsec; 5mV/cm to 2 V/cm calibrated or, d.c. to 60 Mc/s, 50 mV/cm to 20 V/cm. With plug-in unit Model 1085; d.c. to 40 Mc/s (30% down), dual-channel beam switch; rise-time 10 mμsec; 50 mV/cm to 20 V/cm calibrated.

SIGNAL DELAY: 150 mμ sec.

TIME-BASE: 24 calibrated ranges from 0.1 μsec/cm to 5 sec/cm in a 1, 2, 5 series. Continuously variable un-calibrated ranges from 0.1 μsec/cm to 12 sec/cm. Calibrated sweep expansion of x 1 or x 5 on all ranges from 0.1 μsec/cm to 12 sec/cm. D.C. coupled bright-up pulse to CRT grid gives uniform brightness along the trace length and complete fly-back suppression on all ranges. Scan length: 10 cm.

TRIGGER: with plug-in unit Model 1079—five modes of operation, H.F. Sync, Auto, A.C. Fast, A.C. Slow, and D.C. With plug-in unit model 1083—as given above for model 1079 and also with calibrated delayed sweep ranges: 100μsec, 1 msec, 10 msec, 50 msec. With plug-in unit Model 1082—calibrated sweep delay from 2 μsec to 10 sec. Separate trigger selection for main sweep and delaying sweep. Trace may be displayed on delaying sweep. "Lock-out" delay for jitter-free operation.

X AMPLIFIER: with plug-in unit Model 1079—d.c. to 2 Mc/s. 1V/cm to 100 V/cm calibrated. With plug-in unit Model 1083 as given above for Model 1079.

CALIBRATION: Amplitude. A 1 kc/s square-wave calibration source provides eighteen fixed voltage levels from 0.2 mV to 100 V peak-to-peak in a 1, 2, 5 series. Accuracy is within 2 per cent.

Time. Basic calibration ranges are within ± 3 per cent. In addition, a gated 500 (± 2%) Mc/s oscillator can be switched in to provide 2 mμsec intensity modulation dots for accurate measurement of pulse rise-times. Time marker pips are also available at 50 (± 2%) Mc/s.

OUTPUT WAVEFORMS: time-base saw-tooth and gate; vertical-amplifier output; 1 kc/s calibrator square wave.

GENERAL: small neon indicators show the direction in which the CRT beam lies when deflected off the screen. An illuminated rotatable graticle with anti-parallax ruling in centimetre squares with 2 mm base, line divisions and green filter is provided. Fittings for camera attachment.

POWER SUPPLY: Mains. 100 V to 130 V and 200 V to 250 V. Frequency. 50 c/s to 100 c/s. Consumption. 650 W (approximate). Stabilized E.H.T., H.T. and transistor-stabilized d.c. heater supply to drift and hum-sensitive circuits.

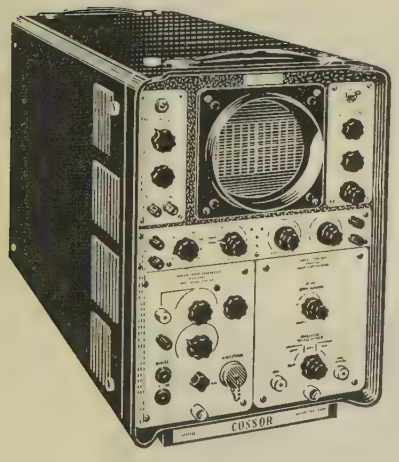
SIZE AND WEIGHT: Height 18½ in. (47.0 cm). Width 13 in. (33.0 cm). Depth 29¼ in. (74.3 cm). Weight with two plug-in units 104 lb (47.3 kg).

ACCESSORIES: Probe Model 1077. Camera Model 1458.

Full details of plug-in units will be supplied on application.

COSSOR INSTRUMENTS

COSSOR INSTRUMENTS LIMITED, COSSOR HO



Around the world in 4 seconds —or 1500 years—these extremes of velocity are attainable by the recording spot of the Cossorscope Model 1076. This high writing speed is one of the many advantages which has earned it acclaim as the most advanced oscilloscope available commercially. The 1076 is yet another significant contribution to research and industry made possible by the imagination and experience of Cossor developmental laboratories.

Cossor manufacture the most extensive range of measuring oscilloscopes available commercially. Please write for details of instrumentation comprising over 50 different aids to industry.

COSSOR
INSTRUMENTS LIMITED

GHBURY GROVE, LONDON, N.5, ENGLAND, TELEPHONE: CANONBURY 1234

ERIE* solid carbon Resistors

The Erie type 15 solid carbon resistor has been introduced to meet the growing demand for sub-miniature components for transistorised equipment.

This new fully moulded resistor is provided with silver clad leads, thus ensuring rapid and effective soldering by whatever method is employed, and like all Erie resistors, it is manufactured to meet the performance standards laid down in DEF.5112.

Still smaller resistors are in course of development.

SPECIFICATION

RESISTANCE RANGE	:	10 ohms to 12 Megohms.
TOLERANCE	:	±5%, ±10% and ±20%.
LENGTH	:	0.200".
DIAMETER	:	0.095".
MAX. PEAK VOLTS	:	50 d.c.
RATING	:	0.1 watt at 70°C.

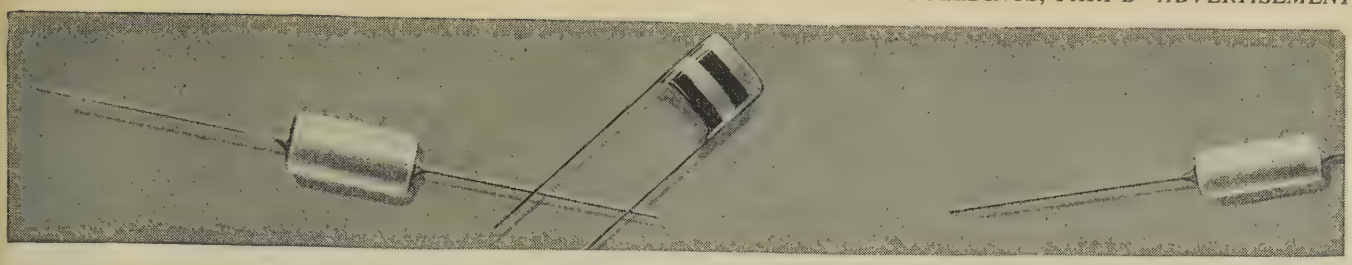
I, HEDDON STREET, LONDON, W.I
Telephone: REGent 6432

FACTORIES
Great Yarmouth and Tunbridge Wells, England: Trenton, Ont., Canada: Erie, Pa., Holly Springs, Miss., and Hawthorne, Cal., U.S.A.

★ ERIE RESISTOR LIMITED

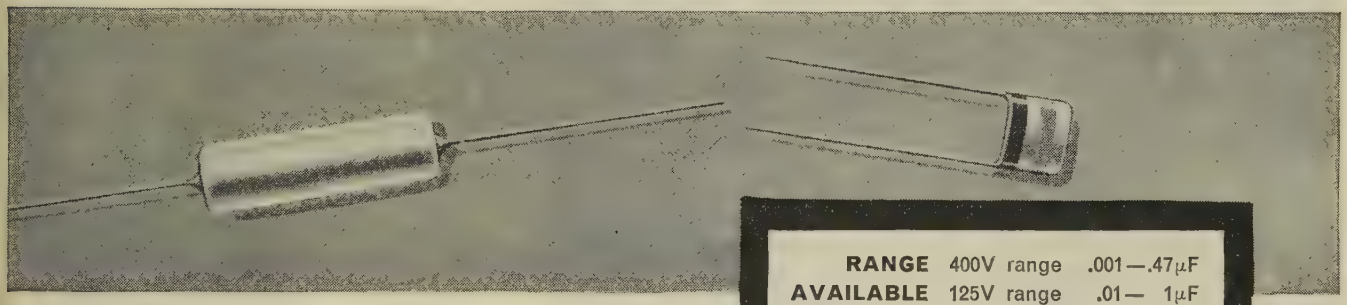
*Registered Trade Mark

Photograph of miniature by courtesy of Victoria & Albert Museum



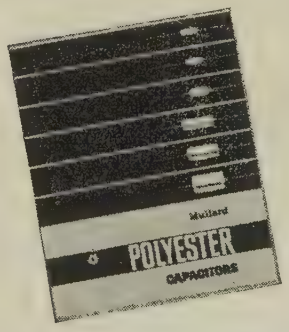
Mullard POLYESTER CAPACITORS...

designed for peak performance with economy...



RANGE AVAILABLE

400V range	.001—.47μF
125V range	.01— 1μF
30V range	.01— .1μF



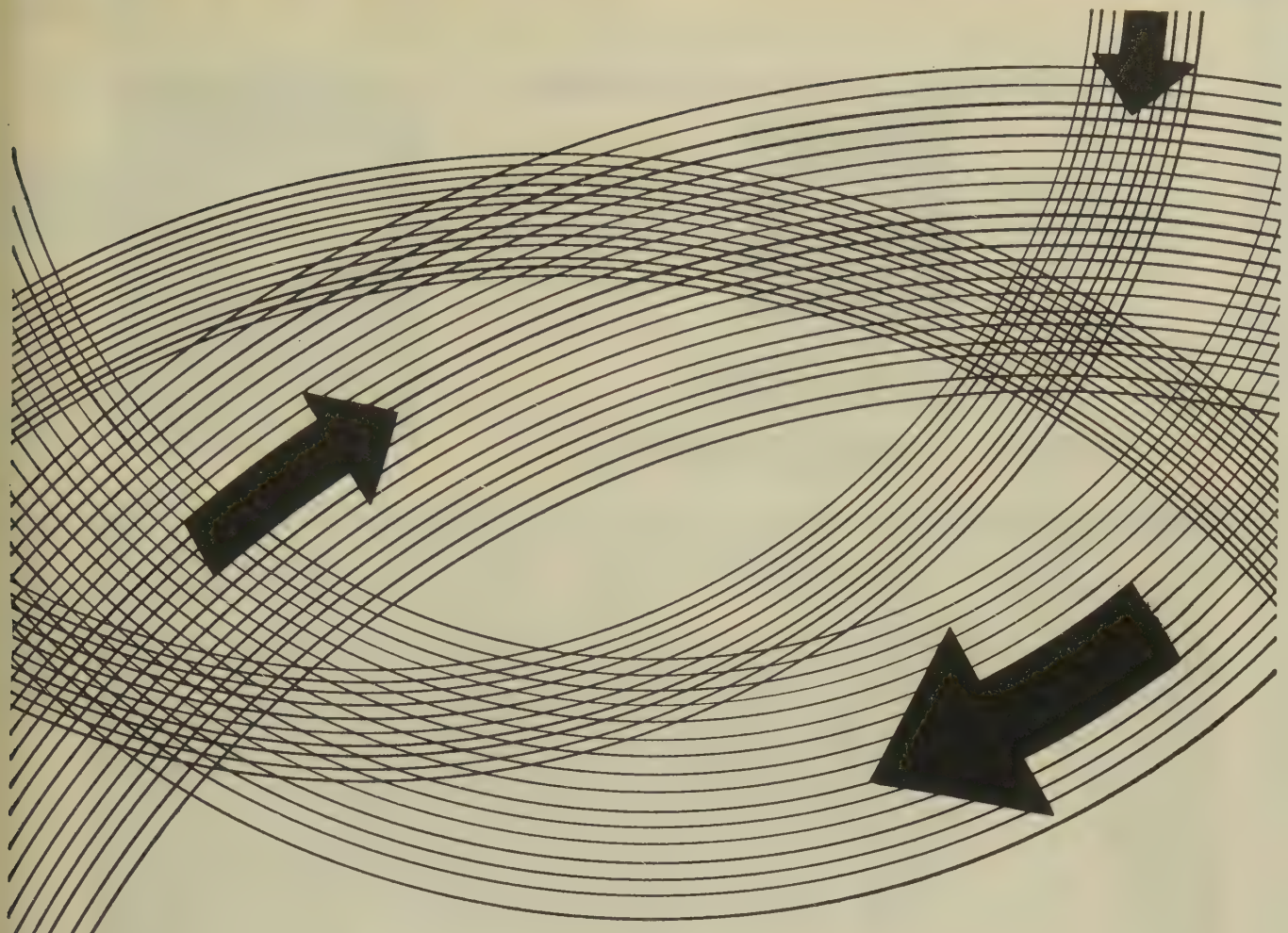
Write to Mullard House for leaflets giving full technical details

- SMALL PHYSICAL SIZE
- HIGH WORKING TEMPERATURES
- CLOSE CAPACITANCE TOLERANCE
- HIGH INSULATION RESISTANCE
- FULLY TROPICALISED
- COST SAVING

Mullard

MULLARD LTD · COMPONENT DIVISION · MULLARD HOUSE · TORRINGTON PLACE · LONDON W.C.1

MC4294



plan with Plannair

Make Plannair a member of your own design team. Many manufacturers requiring temperature control by planned air movement are realising the need to consider this special problem at an early stage—and are calling in Plannair, the air movement specialists, to sit in on their first planning meetings.

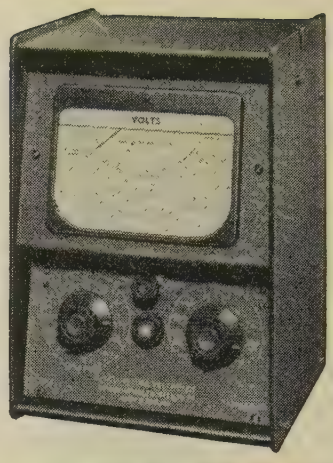
The strength of Plannair lies in the ability of its design engineers to solve complex air movement problems and to design blowers which will provide the right amount of air in the right place for temperature control in specific projects. Size, weight and performance are the prime considerations and these are skilfully balanced in every designed-for-purpose Plannair Blower.

Plannair Limited, Windfield House, Leatherhead, Surrey Tel: Leatherhead 4091

PLANNAIR

PLA64

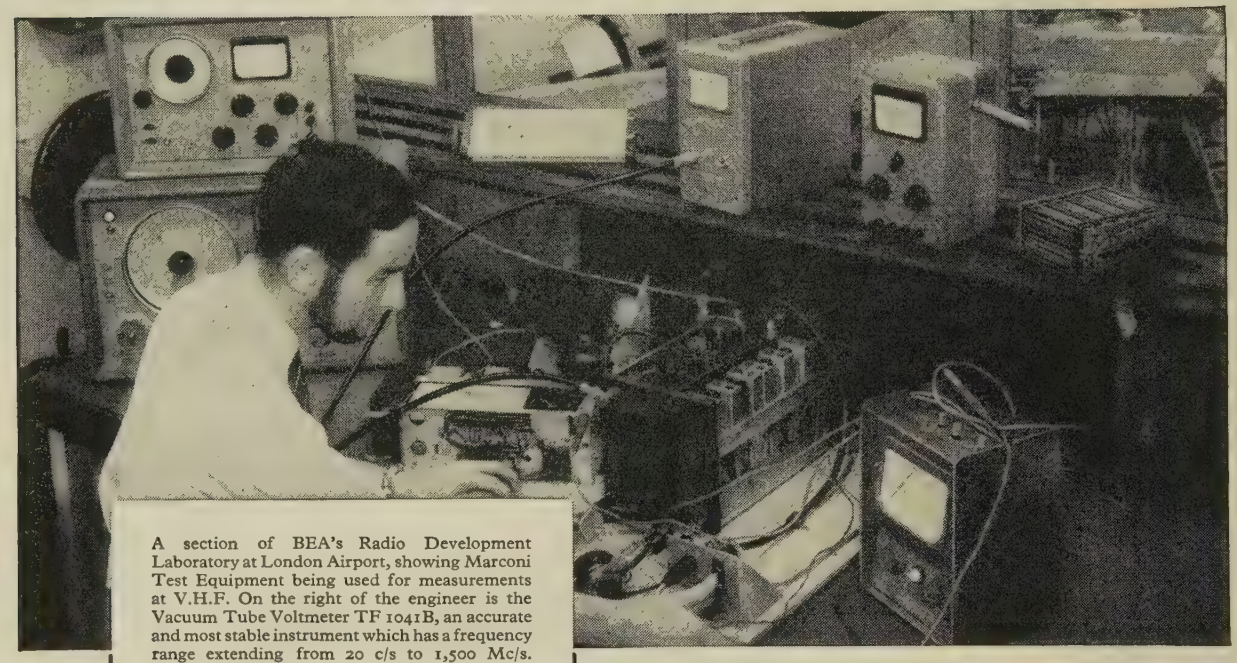
For Versatility and Proven Reliability



BEA

choose

MARCONI TEST EQUIPMENT



A section of BEA's Radio Development Laboratory at London Airport, showing Marconi Test Equipment being used for measurements at V.H.F. On the right of the engineer is the Vacuum Tube Voltmeter TF 1041B, an accurate and most stable instrument which has a frequency range extending from 20 c/s to 1,500 Mc/s. Measures: (i) a.c. up to 300 volts; (ii) d.c. up to 1,000 volts; (iii) resistance, 0.02 ohm to 500 MΩ. A.C. and D.C. multipliers available. For full details, please write for Leaflet K 195.

For the maintenance of high operating standards, the communications and navigational aid systems used in British European Airway's fleet of 100 plus aircraft depend largely on Marconi Test Equipment.

Reliability and ease of manipulation are essential to the hundreds of settings and measurements made each day in the Overhaul Workshop and the Development Laboratory at London Airport. The many applications and techniques of modern airborne equipment necessitate accurate measurement of voltage and power at a wide range of levels, and at frequencies varying from d.c. to hundreds of megacycles. Exact measurement of inductance, capacitance and impedance is also essential and, of course, precision signal sources are constantly in use. It is not surprising, therefore, that Marconi Test Equipment, largely owing to its versatility and proven reliability, comprises 59% of BEA's total requirement.

MARCONI INSTRUMENTS

THE INTERNATIONAL CHOICE FOR ELECTRONIC MEASUREMENT

AM & FM SIGNAL GENERATORS · AUDIO & VIDEO OSCILLATORS · FREQUENCY METERS · VOLTMETERS · POWER METERS · DISTORTION METERS
TRANSMISSION MONITORS · DEVIATION METERS · OSCILLOSCOPES · SPECTRUM & RESPONSE ANALYSERS · Q METERS & BRIDGES

London and the South:
English Electric House, Strand, W.C.2
Telephone : COVent Garden 1234

Midlands:
Marconi House, 24 The Parade, Leamington Spa
Telephone : 1408

North:
23/25 Station Square, Harrogate
Telephone : 67455

Export Department Marconi Instruments Ltd., St. Albans, Herts, England. Telephone : St. Albans 59292

REPRESENTATION IN 68 COUNTRIES

TC 195

AEI 100 AND 80 AMP SILICON-CONTROLLED RECTIFIERS

FOR MOTOR CONTROL, HIGH-POWER STATIC SWITCHING, SERVO SYSTEMS, HIGH-POWER INVERTORS, AND REGULATED D.C. POWER SUPPLIES.

ADVANCE INFORMATION

OUTSTANDING FEATURES

★ HIGH POWER (TENS OF KILOWATTS) CAN BE CONTROLLED BY LESS THAN 2 WATTS ★ LOW FORWARD VOLTAGE DROP—THUS HIGH ELECTRICAL EFFICIENCY ★ RUGGED ENCAPSULATION ★ HERMETIC SEAL ★ HIGH-EFFICIENCY HEAT SINK MOUNTING

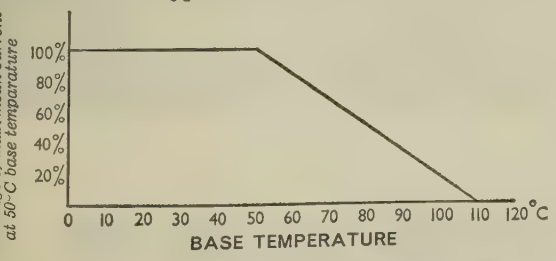
100-AMP TYPES CR100-	21	51	71	101	151	201	251	301
Peak Forward Working Voltage (Volts)	25	50	75	100	150	200	250	300
Peak Reverse Voltage (Volts)	25	50	75	100	150	200	250	300
Max. Mean Forward Current (Amps)*	100	100	100	100	100	100	100	100
Max. Recurrent Peak Forward Current (Amps)	500	500	500	500	500	500	500	500
Max. Forward Voltage Drop at 350 Amps Peak (Volts)	2	2	2	2	2	2	2	2
Max. Trigger Firing Current Required (Amps)	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
Max. Trigger Firing Voltage Required (Volts)	4	4	4	4	4	4	4	4
Max. Peak Trigger Current (Amps)	2	2	2	2	2	2	2	2
Max. Mean Trigger Current (Amps)	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
Peak Forward Trigger Voltage —anode + ve (Volts)	10	10	10	10	10	10	10	10
Peak Forward Trigger Voltage —anode —ve (Volts)	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
Peak Reverse Trigger Voltage	5	5	5	5	5	5	5	5

80-AMP TYPES CR80-	21	51	71	101	151	201	251	301
Peak Forward Working Voltage (Volts)	25	50	75	100	150	200	250	300
Peak Reverse Voltage (Volts)	25	50	75	100	150	200	250	300
Max. Mean Forward Current (Amps)*	80	80	80	80	80	80	80	80
Max. Recurrent Peak Forward Current (Amps)	500	500	500	500	500	500	500	500
Max. Forward Voltage Drop at 350 Amps Peak (Volts)	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5
Max. Trigger Firing Current Required (Amps)	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
Max. Trigger Firing Voltage Required (Volts)	4	4	4	4	4	4	4	4
Max. Peak Trigger Current (Amps)	2	2	2	2	2	2	2	2
Max. Mean Trigger Current (Amps)	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
Peak Forward Trigger Voltage—anode + ve (Volts)	10	10	10	10	10	10	10	10
Peak Forward Trigger Voltage—anode—ve (Volts)	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
Peak Reverse Trigger Voltage	5	5	5	5	5	5	5	5

*At a conduction angle of 180° and base temperature of 50°C.

Frequency: The ratings given apply without derating to sine wave operation up to 3 Kc/s.

Temperature: The provisional curve for derating with temperature is shown below.

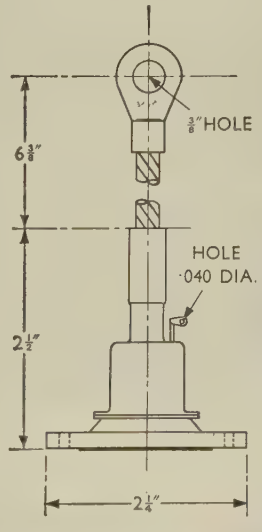


FEATURES

Encapsulation All AEI Silicon-Controlled Rectifiers are hermetically sealed in metal glass enclosures backfilled with dry inert gas to ensure long life in all environments.

Mounting The CR100 and CR80 are mounted on a substantial nickel plated copper base providing efficient thermal and electrical contact. A fixing ring is provided with each controlled rectifier.

Dimensions for all types are as shown on the right.



PLEASE DON'T TEAR OUT THIS ADVERTISEMENT—

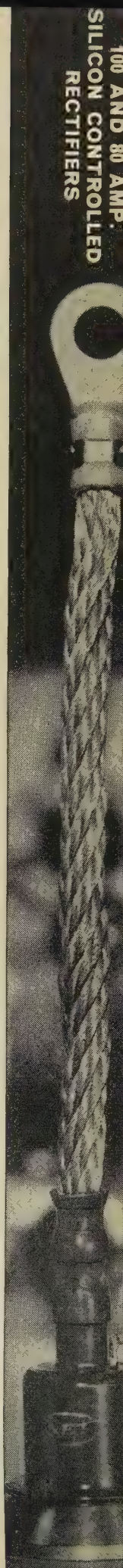
Others will want to see it. If you would like to have full details of these, or any other AEI electronic components for that matter, all you have to do is write to AEI at the address below. Data sheets with detailed information will be sent to you without charge.

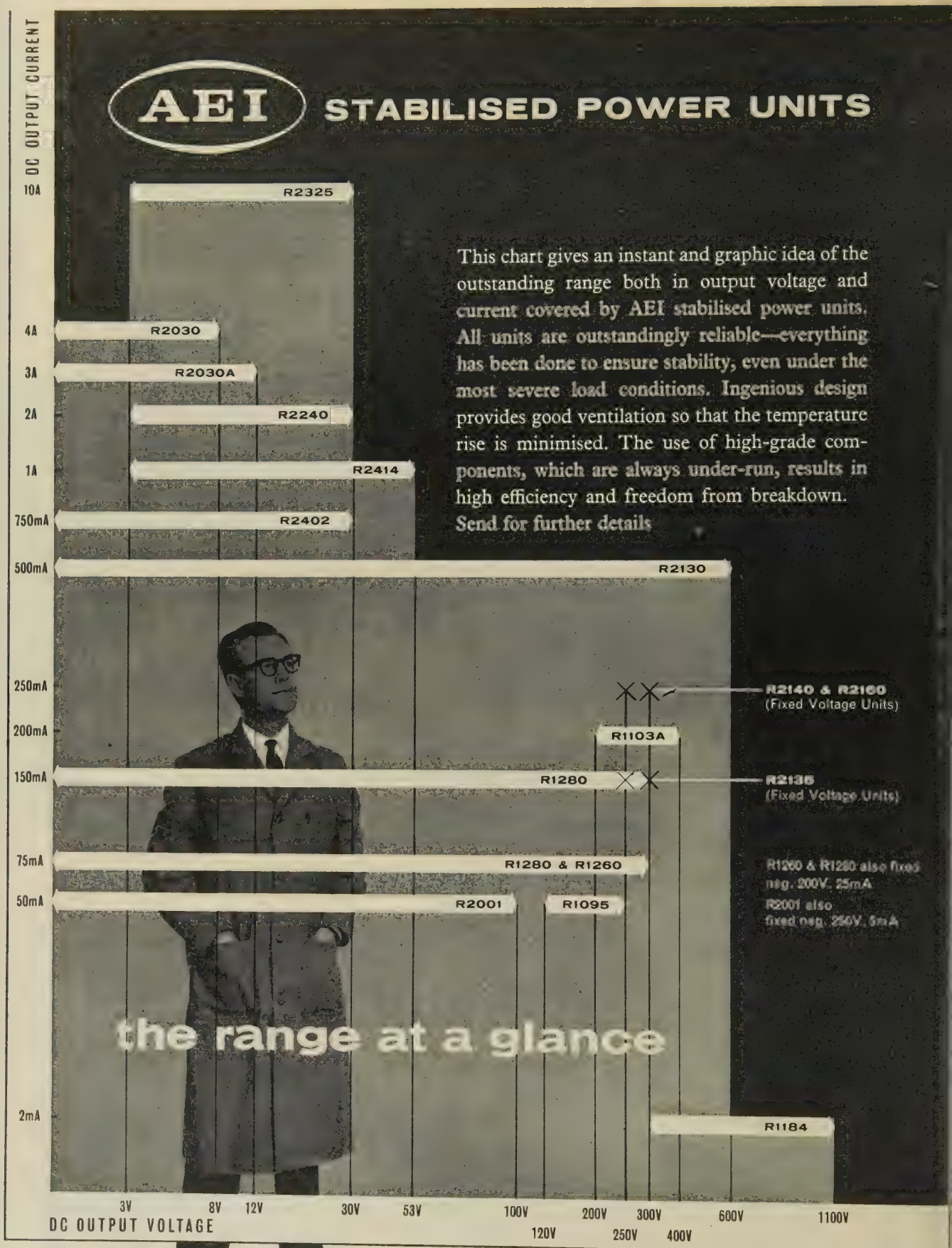
DON'T DELAY—*write now and be sure of receiving latest news of these and other AEI components as they become available.*

AEI **Associated Electrical Industries Limited**

ELECTRONIC APPARATUS DIVISION, VALVE & SEMI-CONDUCTOR SALES DEPT.
CARHOLME ROAD · LINCOLN · TEL: LINCOLN 26435

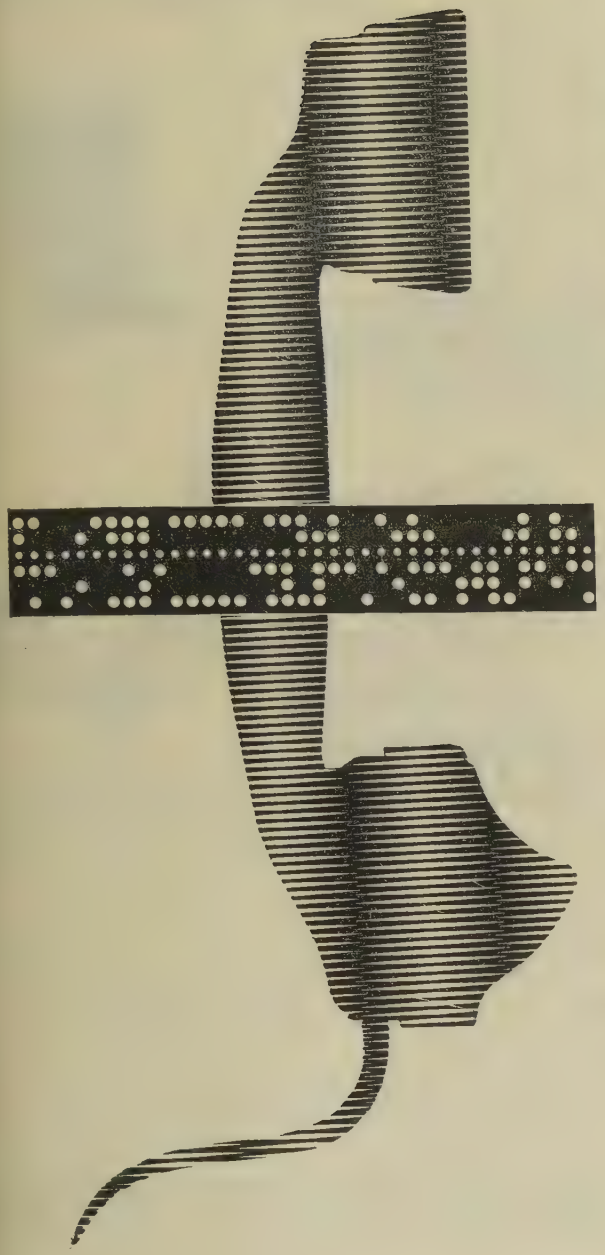
A5671





This chart gives an instant and graphic idea of the outstanding range both in output voltage and current covered by AEI stabilised power units. All units are outstandingly reliable—everything has been done to ensure stability, even under the most severe load conditions. Ingenious design provides good ventilation so that the temperature rise is minimised. The use of high-grade components, which are always under-run, results in high efficiency and freedom from breakdown. Send for further details

Associated Electrical Industries Ltd
Radio & Electronic Components Division
PD 17, 155 Charing Cross Road, London W.C.2
Telephone: GERrard 9797 Telegrams: Sieswan Westcent Lonc
17/1



FIRST RAILWAY MICROWAVE RADIO TELEPHONE SYSTEM IN BRITAIN

300 Channels between Newcastle and York

British Railways first microwave multichannel system from Newcastle to York via Darlington will have a 300 telephone channel capacity. The system allows for channels to be dropped off at intermediate points and can accommodate high speed data transmission.

MARCONI

COMPLETE COMMUNICATION SYSTEMS
SURVEYED · PLANNED · INSTALLED · MAINTAINED

COMMUNICATIONS DIVISION, MARCONI'S WIRELESS TELEGRAPH COMPANY LIMITED, CHELMSFORD ESSEX, ENGLAND.

H 3

A **MARCONIDATA**

DATA COMMUNICATION SYSTEM
DOESN'T EXIST
UNTIL WE BUILD IT FOR YOU

MARCONIDATA is backed by a name made famous in every country of the world for telecommunications practice and technique.

MARCONIDATA the product of the largest research and development organization of its type in Europe, is a completely flexible system composed of standard sub-units which can be arranged to provide accurate data communication between any number of points over any distance.

MARCONI

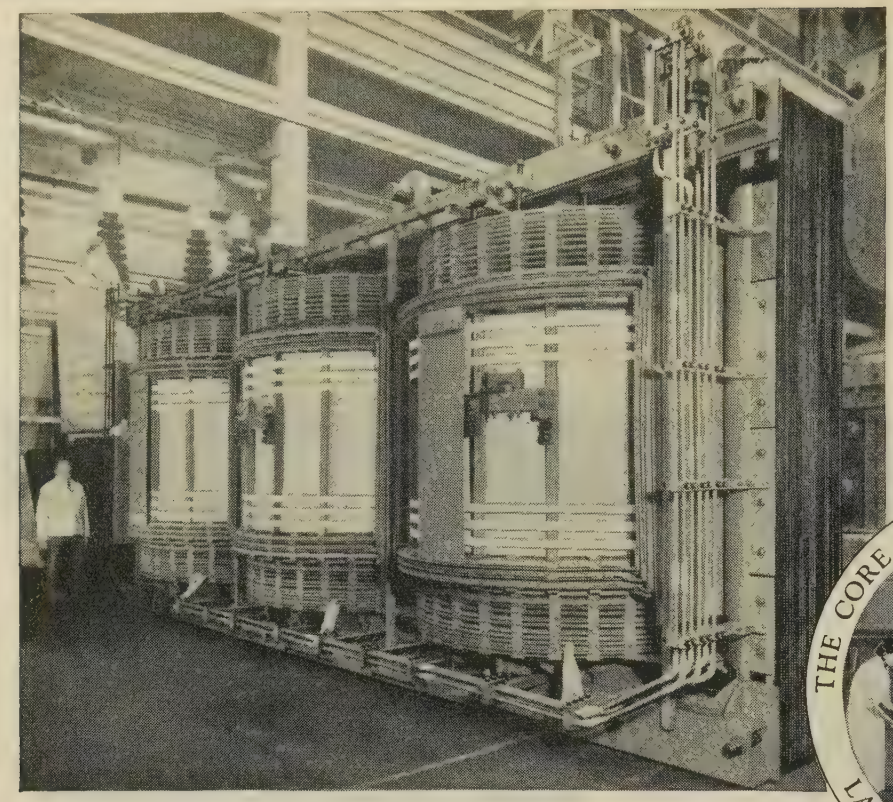
COMPLETE COMMUNICATIONS SYSTEMS

SURVEYED · PLANNED · INSTALLED · MAINTAINED

FAST ACCURATE DATA COMMUNICATION SYSTEMS

MARCONI'S WIRELESS TELEGRAPH COMPANY LIMITED, CHELMSFORD, ESSEX, ENGLAND

B4



LAMINATIONS

of all types, in all sizes and in all grades of material.

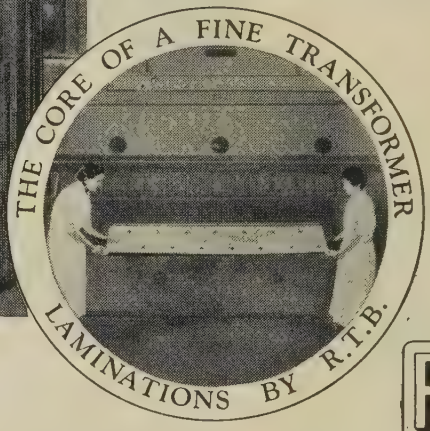
FERROSIL

hot-rolled and cold-reduced electrical sheet and strip, and hot-rolled transformer sheet.

ALPHASIL

cold-reduced oriented transformer sheet and strip.

One of many three-phase, 50 cycle, 120,000 kVA, 275/132 kV Auto-transformers supplied by the BTH Company for the British Super Grid.



RTB

RICHARD THOMAS & BALDWINS LTD.

Enquiries for sheet and strip to be addressed to RICHARD THOMAS & BALDWINS (SALES) LIMITED, WILDEN, STOURPORT-ON-SEVERN, WORCS.
Enquiries for laminations to be forwarded to RICHARD THOMAS & BALDWINS LIMITED, COOKLEY WORKS, BRIERLEY HILL, STAFFS.

---

for a high quality self-fluxing enamelled wire

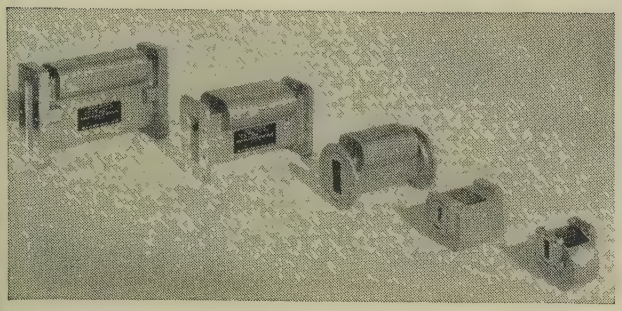
Superfine Lewcosol

is the unanimous choice

In response to the growing demand for high quality solderable superfine enamelled wires the range of Lewcosol wires has now been extended down to ·001 inch. Manufactured under strictly controlled conditions on newly developed plant, these wires will find immediate application where size, solderability, and a consistently high quality are essential requirements.

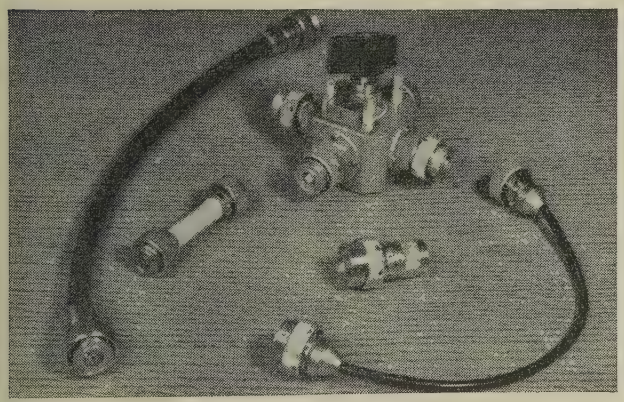
THE LONDON ELECTRIC WIRE COMPANY AND SMITHS LIMITED
LEYTON · LONDON · E.10

Ferrite Isolators



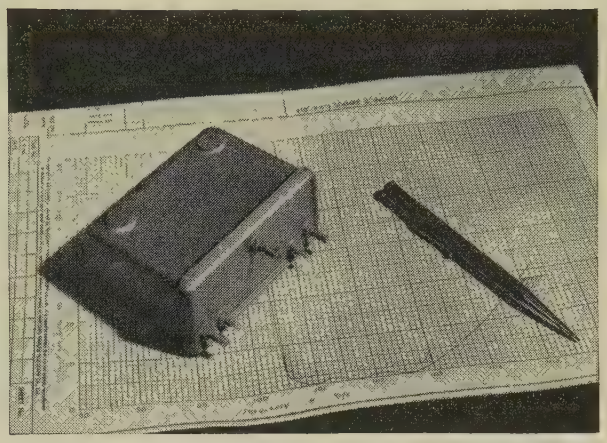
Existing Marconi designs cover most of the frequencies from 1.8 Kmc/s to 10 Kmc/s, and enable the following performance characteristics to be obtained :
Reverse losses of 45 dB or greater VSWRs 1.03.
Forward losses of 0.5 dB or less.
A wide range of ferrite devices, including circulators, is available; if your requirements fall outside our existing range we shall be pleased to consider a design to suit you.

50 ohm Coaxial Connectors



Designed to combine low VSWR with positive mechanical connection at 5000 mc/s, this connector is unique in its use for laboratory measurements and for any application calling for a stable and reliable match. Alternative arrangements are available for flexible cables or panel mounting use. The connectors are normally female but can be converted to male by the simple addition of a locating ring and connecting pin. The design incorporates a sealing ring for waterproofing the connecting joint.

Miniature Crystal Filters



Crystal filters operating at 100 kc/s with bandpass characteristics.
Pairs of upper and lower sideband selection filters for SSB applications.
IF filters at 455 kc/s.
Marconi's are constantly engaged in the development of new crystal filter techniques, and it is impossible to specify the large number of filter designs available. The Company's unique experience in this field enables us to offer advice and technical assistance with filter problems of any kind.

MARCONI

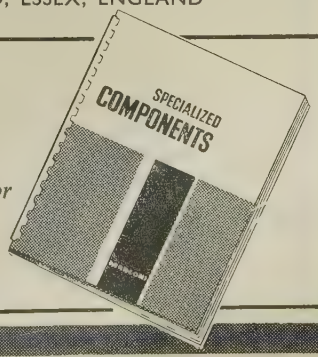
SPECIALIZED RADIO COMPONENTS

Write for details of crystal filters and other specialized components to :

SPECIALIZED COMPONENTS GROUP

MARCONI'S WIRELESS TELEGRAPH COMPANY LIMITED, CHELMSFORD, ESSEX, ENGLAND

A preliminary catalogue is available listing specialized components for immediate delivery.



W3

PAPERS FOR THE PROCEEDINGS

Handbook for Authors

Anyone who is thinking of submitting a Paper to The Institution should apply to the Secretary for a copy of the **Handbook for Authors**. The price is 3s. (post free), but a copy will be supplied free of charge if the application is accompanied by a summary of the Paper. The following are some of the main points considered in the Handbook.

Acceptability

To be acceptable, a Paper should normally contribute to the advancement of electrical science or technology. The Institution does not accept Papers which have been published elsewhere.

Length

No Paper should occupy more than 10 pages in the *Proceedings*. Authors can generally keep well within this limit. For example, the average Paper published in 1960 consisted of 6 000 words (5 pages) and, with its illustrations and mathematics, occupied a total of 8 pages.

Summary

An essential part of a Paper is the **Summary**, which should not exceed 200 words.

Text

The **Text** should begin with sufficient introductory matter to enable the Paper to be understood without undue reference to other publications.

The Text should include no more mathematics than is essential. Extended mathematical treatment and lengthy digressions—if they must be included—should be put in **Appendices.**

Proprietary articles should not be mentioned by name unless this is unavoidable.

The rationalized M.K.S. system of units is preferred.

Acknowledgments

Assistance in the preparation of the Paper, and sources of information, should be acknowledged. References to manufacturers should be made only under **Acknowledgments.**

References

Bibliographical **References** should be numbered and listed in a special section, and indicated in the Text by means of 'indices'.

Numbering

The Text should be appropriately sectionalized, the sections and their subdivisions being numbered according to the 'decimal' system. Acknowledgments, References and Appendices should be numbered as though they were sections of the Text.

Typing

Typing should be on one side of the paper only, with double spacing between lines and a 1½ inch margin on the left. Besides the original typescript, two carbon copies are required by The Institution.

Advice on the typing of mathematics is given in the **Handbook for Authors,** which includes a facsimile of a typewritten page containing mathematics.

Illustrations should not be drawn or pasted on the typewritten pages. They are of no use to the printer, but he does need a complete list of captions, again with double spacing. The list should be attached to the typescript.

Drawings

Three sets of drawings, which may be in the form of dye-line prints, should accompany the typescript. Tracings, which will be required later, should be in indian ink with the lettering in pencil. The reduction in the size of the drawings, and therefore the size of lettering required, is not settled before the Paper has been accepted.

Dispatch

The typescript and illustrations should be packed flat, not rolled, and addressed to *The Secretary, The Institution of Electrical Engineers, Savoy Place, London, W.C.2.*

HIGH GAIN QUADRANT AERIALS

FOR BANDS I, II, AND III

* *Simple construction and simple to erect*
* *Unit construction with distribution feeder incorporated in each stack*
* *Low wind loading*
* *Radiation pattern can be omnidirectional or directional*
* *Reliable — no mechanically stressed insulators to break down*
* *Parallel-working transmitters can be connected*
* *High Gain — 1·2 times per stack*
* *For pole or lattice steel mast mounting*

MARCONI

COMPLETE SOUND AND TELEVISION SYSTEMS

MARCONI'S WIRELESS TELEGRAPH COMPANY LIMITED · CHELMSFORD · ESSEX · ENGLAND

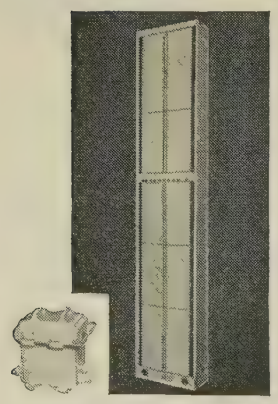
Cable carrier systems by A.T.E.

The A.T.E. Range of Cable Carrier Systems features equipment for large and small capacity routes. All systems meet the internationally recognised C.C.I.T.T. requirements for trunk circuits, are of high quality and advanced design. Write for further details to:—

AUTOMATIC TELEPHONE & ELECTRIC CO LTD

Strowger House, Arundel Street, London, W.C.2. Phone: TEMple Bar 9262

A.T.E. TRANSMISSION EQUIPMENT TYPE CM FOR LINE, CABLE AND RADIO SYSTEMS

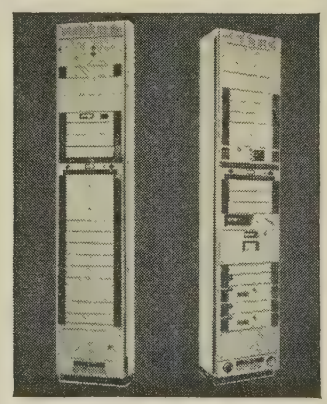


Right: Terminal Repeater
Left: Intermediate Buried Repeater

C300A Small Core Coaxial System

- Fully Transistorised.
- 300 channel system for small core coaxial cables.
- Intermediate power fed repeaters in sealed buried boxes.
- Automatic pilot regulation, suitable for buried or aerial cable.
- Power feeding stations may be up to 60 miles apart.
- Inbuilt maintenance and fault location facilities.

For details see Bulletin TEB 3202

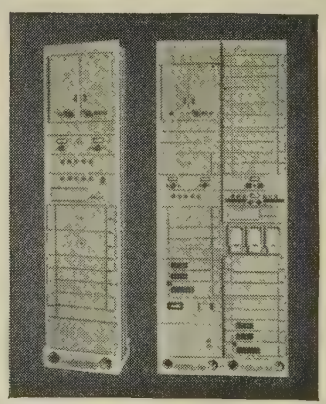


Left: Terminal Repeater (Receive)
Right: Terminal Repeater (Transmit)

C960A 4 Mc/s Coaxial System

- Up to 960 high-grade telephone circuits on each pair of conventional coaxial tubes.
- Power fed dependent repeaters at 6 mile spacing.
- Main power feed stations up to 100 miles apart.
- Comprehensive maintenance and test facilities.
- Conforms to C.C.I.T.T. recommendations.

For details see Bulletin TEB 1411

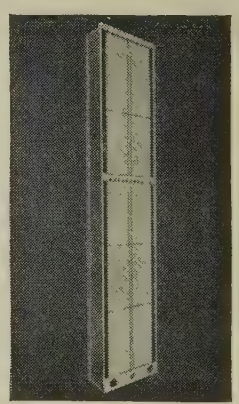


Left: Dependent Repeater—6 ft.
Right: Terminal Repeater—9 ft.

CX12A 12.5 Mc/s Coaxial System

- Up to 2,700 high grade telephone circuits or transmission of mixed traffic on a pair of conventional .375 inch dia. coaxial tubes.
- Conforms to C.C.I.T.T. recommendations and G.P.O. specifications.
- Dependent repeaters power fed from terminal equipment, with automatic transfer to local mains supply in case of failure.
- Comprehensive maintenance and test facilities.

For details see Bulletin TEB 1417



Terminal Rackside

C12G Cable Carrier System

- Fully Transistorised.
- 12 Channels on a single cable pair. (6-54 Kc/s and 60-108 Kc/s 'go' and 'return').
- Automatic pilot regulation, suitable for aerial or buried cables.
- Straight or 'frogging' repeaters.
- Terminal for 2 complete systems with signalling and frequency generating equipment on one 9 ft. rackside.

For details see Bulletin TEB 3201

Quartz Crystal Ovens

STABLE TEMPERATURE ENSURING MAXIMUM FREQUENCY STABILITY

Switching differential 0·0014°C.
No thermostat
No thermometer switch

Orthodox crystal ovens, using thermostats or thermometer switches, are available for applications where wider temperature variations are acceptable.

This symbol has been adopted by the Marconi Specialized Components Group. The Marconi Company undertakes the design and manufacture of specialized components only when no suitable alternative is available; and in almost every case, Marconi specialized components are designed to more exacting standards and built to closer tolerances than any similar components. A preliminary catalogue is available, listing ferrite isolators and circulators, coaxial connectors, attenuators, terminations and switches; waveguide filters terminations, and bends; crystal filters and ovens.

MARCONI

SPECIALIZED RADIO COMPONENTS

Write for details of crystal ovens and other specialized components in the Marconi range, and address your enquiries to:

SPECIALIZED COMPONENTS GROUP

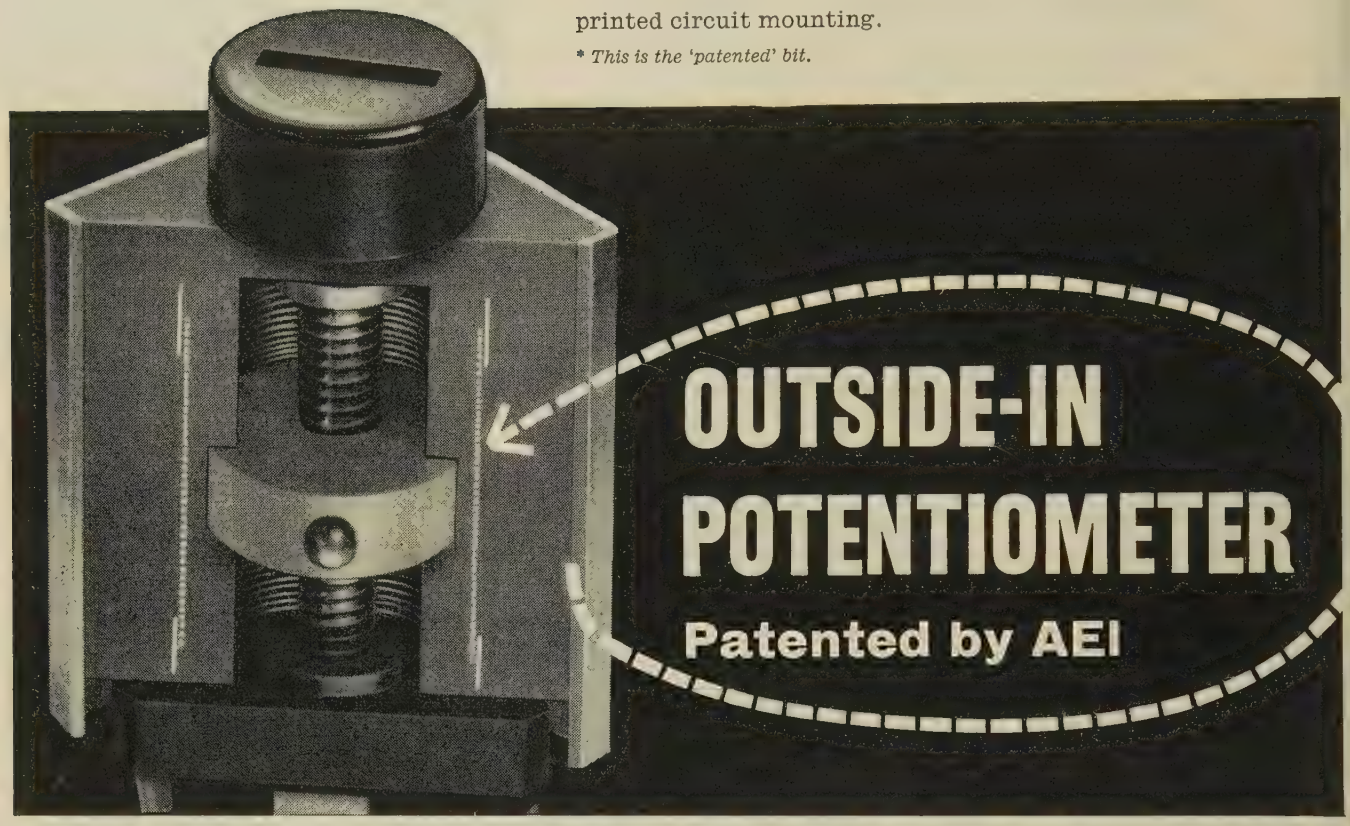
MARCONI'S WIRELESS TELEGRAPH COMPANY LTD.,
CHELMSFORD, ESSEX, ENGLAND

A WHOLE HOST OF ADVANTAGES spring from the revolutionary design and construction of the new AEI 'Minitrim' Miniature Potentiometer. The secret is that the element, instead of being on the outside, is encapsulated* in epoxy resin.

So—the element cannot move under the action of the brush; bunched turns are practically eliminated; oxidisation of the track face is minimised; it is almost impossible for moisture or dust to enter; acceleration or vibration have no effect; heat dissipation is high because of the proximity of element and aluminium can.

The Minitrim is available for individual, stack, or printed circuit mounting.

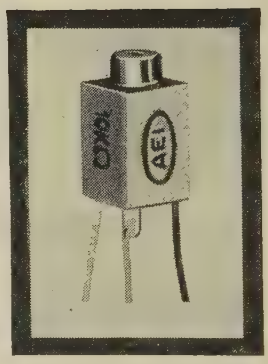
* *This is the 'patented' bit.*



DATA

Resistance values	50kΩ to 10kΩ†
Resistance tolerance	±20%
Insulation resistance	1000MΩ at 500V D.C.
Max. operating voltage	500V
Wattage rating	1W at 20°C ambient 0.5W at 60°C ambient
Temperature range	—40°C to 125°C
Number of operating turns	20 approx.

† Values up to 20kΩ will be available shortly



DIMENSIONS

(designed on 0.5 in. module)
Length 0.8 in. Cross section 0.4 in. x 0.4 in.
Weight 3 gms

AEI ELECTRONIC COMPONENTS

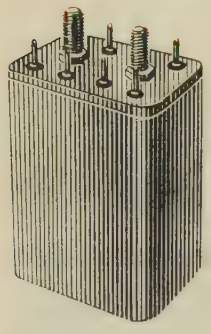
Send now for full details of the 'MINITRIM' MINIATURE POTENTIOMETER

Associated Electrical Industries Limited
Radio & Electronic Components Division
155 Charing Cross Road, London WC2
Telephone: GERrard 9797

TA17/104

INDEX OF ADVERTISERS

Plessey relays for the electrical industry



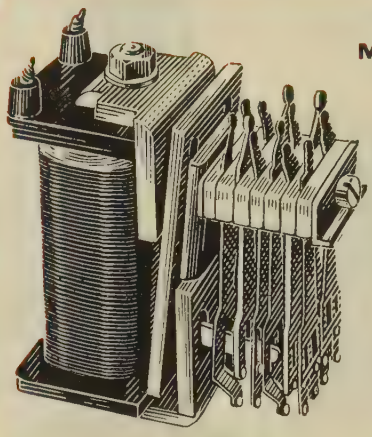
MINIATURE TYPE CA

Fully Type Approved to RCS 165 and 166.

Light and medium duty types have two changeover contacts.

Heavy duty types single make or break.

Available sealed, or unsealed with dustcover.



MINIATURE TYPE CB

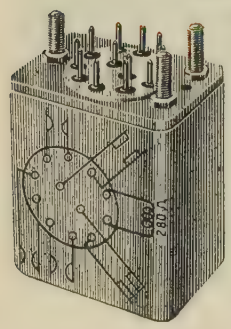
Based on Type CA with heavier magnetic circuit.

All versions fully Type Approved.

Available with up to twelve contact springs.

Twinned platinum contacts on light duty versions.

Available sealed, or unsealed with dustcover.



MINIATURE TYPE CC

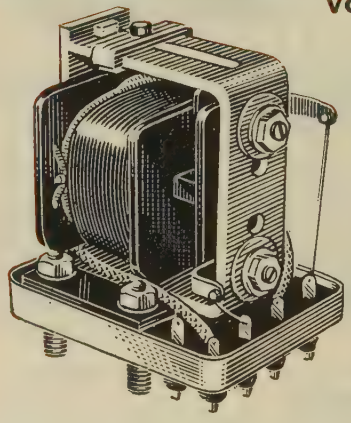
Meets RCS 165 and MIL-R-5757 Specifications.

Based on Type CA magnetic circuit, providing up to four light or medium-duty contact sets.

Available sealed, or unsealed with dustcover.

Printed circuit versions available.

Contact loadings up to 10A d.c., non-inductive, can be arranged for Types CA, CB, and CC.



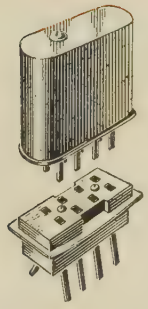
VOLTAGE REGULATING TYPE XC 269

Fully Approved to S.R.D.E. Spec. 166/1.

Temperature compensated in range −40°C to +85°C.

Available for 6, 12 or 24 V operation.

Typical changeover voltage differential 1 V in 24 V.



SUB-MINIATURE TYPE CE

Occupies less than ¼ square inch of chassis area.

Two changeover contacts rated 0.25A at 28 V d.c.

Insulation proof against 500 V a.c., r.m.s. between coil and contact stack.

−55°C to +100°C operational temperature range.

Plessey

THE PLESSEY COMPANY LIMITED
(Relays and Control Systems Unit)

Eddes House, Eastern Avenue West, Romford, Essex
Telephone: Seven Kings 6050

Overseas Sales Organisation: Plessey International Limited, Ilford, Essex. Telephone: Ilford 3040

RCS 2

THE PROCEEDINGS OF
THE INSTITUTION OF ELECTRICAL ENGINEERS

EDITED UNDER THE SUPERINTENDENCE OF W. K. BRASHER, C.B.E., M.A., M.I.E.E., SECRETARY

VOL. 108. PART B. NO. 41. SEPTEMBER 1961

621.396.67

The Institution of Electrical Engineers
Paper No. 3663 E
Sept. 1961

AERIALS

A Review of Progress

By H. PAGE, M.Sc., Member.

(1) INTRODUCTION

The function of the transmitting aerial in any radio system is to launch electromagnetic waves into space. It may be regarded as a leaky waveguide, the degree of guidance in launching the waves being controlled by the configuration of the aerial. The most important characteristics are the radiation pattern, i.e. the intensity of the radiation as a function of direction, and the impedance. The shape of the radiation pattern is a measure of the success achieved in concentrating the energy in the required directions, and the impedance determines the matching to the source. Correspondingly, a receiving aerial is required to act as a collector of electromagnetic waves flowing in space. Since an aerial and the associated propagation medium generally constitute a linear passive network, it follows that the radiation pattern and the impedance are the same whether the aerial is used for transmission or reception.* In any specific problem we may therefore consider whichever case is the easier to analyse.

The dimensions of the aerial enter into the expressions for the radiation pattern and impedance only as fractions of the wavelength,† so that the same theoretical design methods can be used in all wavebands. At the shorter wavelengths, however, the physical dimensions of a given type of radiating element are smaller, and techniques may then be used that would otherwise be impracticable. For instance, for long wavelengths the aerial usually consists of one or more conductors separately connected to the generator through transmission lines. For decimetric and shorter wavelengths, on the other hand, it is more convenient to use reflecting surfaces or refracting media to shape the radiation pattern, these being techniques akin to those used in optics.

It is therefore convenient to divide this review according to the waveband in which the design techniques described find important application. There is unfortunately no generally accepted method of designating the bands, and the nomenclature used here is that recommended at the ninth plenary assembly of the Comité Consultatif International Radio at Los Angeles in 1959, namely:

Frequency range	*Wavelength range*	*Metric subdivision*
3–30 kc/s	100–10 km	Myriametric waves
30–300 kc/s	10–1 km	Kilometric waves
300–3 000 kc/s	1 000–100 m	Hectometric waves
3–30 Mc/s	100–10 m	Decametric waves
30–300 Mc/s	10–1 m	Metric waves
300–3 000 Mc/s	100–10 cm	Decimetric waves
3–30 Gc/s	10–1 cm	Centimetric waves

This is the first review of aerials to appear in the *Proceedings* of The Institution, and it is therefore appropriate to describe the developments that have taken place from the earliest days; this means that only the major trends can be covered, and these somewhat superficially. However, survey papers dealing more fully with specific aspects of the subject are referred to, and the reader may use them not only to fill in some of the detail, but also as sources of further references.

(2) HISTORICAL SURVEY

The first demonstration of the radiation of electromagnetic waves is generally regarded as that given by Hertz in 1888, although other workers were experimenting along the same lines at, or before, that time. Hertz' aerial consisted of two spheres 30 cm in diameter, spaced 1 m apart; they were connected by thin wires to a central gap across which sparks were induced by means of a large induction coil. The term 'doublet' is now used to describe an aerial consisting of two spheres of large capacitance spaced a small fraction of the wavelength apart and joined by a thin wire in which a current is made to flow; although no longer used as a practical aerial it is a most useful concept in the development of the theory. Hertz calculated the field of, and the power radiated by, such an aerial.

Hertz' later experiments were directed primarily to investigating the nature of electromagnetic radiation, and included a demonstration of the focusing properties of a parabolic cylinder. This work inspired Righi, Bose, Marconi and Lodge to pursue the correspondence between electromagnetic and optical phenomena; an account of this work has been given by Ramsay.[1]

* In some special cases mentioned in Section 7.5 the system is not linear, and the reciprocity principle does not then apply.

† Here we assume the materials used are lossless. This is virtually true in practice, except when the ground is one of the terminals.

Mr. Page is in the Research Department of the British Broadcasting Corporation.

It is remarkable to find that, in addition to investigations on radiation and polarization phenomena, these workers were experimenting also with parabolic mirrors, radiating horns and lenses, a field which has been seriously reinvestigated only during the past 20 years.

Between 1910 and 1916 the propagation of electromagnetic waves along dielectric rods was the subject of study by a number of workers,[2] but again a long period was to elapse before interest in this subject was revived as the result of improved experimental facilities for short wavelengths.

The principle of modulating the generator to transmit intelligence was demonstrated by Marconi in 1896 on a wavelength of 25 cm, and his success focused attention on the potential importance of electromagnetic radiation for communication purposes. For a considerable period thereafter the emphasis was on the exploitation of the new medium to achieve larger ranges, rather than on the scientific importance of the phenomena involved. The first aerials took the form of Hertzian doublets, but Marconi soon appreciated that the 'grounded aerial' (in which the ground was used as one of the terminals) was simpler and just as effective. Higher aerials thus became practicable, and the longer wavelengths of the radiation from these large structures were found to give increased range of communication. Aerials were built with a large number of wires in parallel to cater for the high voltages developed by the powerful spark transmitters.[3] For instance, the first aerial at the Poldhu transmitting station in England was made of 2000 wires in the shape of a cylinder 200 ft high and 150 ft in diameter; it was this station which, in 1901, made history by establishing communication with St. John's, Newfoundland, an achievement far in advance of the scientific thought of the day. Subsequent progress was limited primarily by the availability of suitable generating equipment, but the First World War stimulated rapid developments. In particular, the introduction of the valve made it possible to generate, and to amplify, continuous-wave oscillations.

The inauguration of experimental broadcasting on hectometric wavelengths in 1922 provided a further fillip to development. At first single-wire transmitting aerials were used, but as the output power of transmitters increased it became necessary to revert to Marconi's idea of using multi-wire aerials of large capacitance, in order to avoid excessive voltages. Later it became apparent that it was not the strength of the signal but selective fading (accompanied by distortion of the programme) that was imposing a limit on the useful range, and that to overcome this restriction it was necessary to reduce upward radiation towards the ionosphere. The foundation for the next step had been laid by Pierce,[4] who in 1916 had calculated the radiation pattern of a vertical aerial carrying a sinusoidally distributed current. In 1934, Ballantine[5] went on to show that the maximum service area on hectometric wavelengths is achieved by using aerials $0{\cdot}5$–$0{\cdot}6\lambda$ in height, the optimum height depending on the wavelength and the ground conductivity. Such aerials, which nowadays usually take the form of insulated masts, are described as anti-fading. Subsequent work, described in Section 3, has led to a fuller understanding of the performance of such aerials.

In 1916, Marconi and Franklin started a series of experiments which was to have far-reaching effects; they successfully beamed radiation on a wavelength of 3 m by means of a reflector, composed of tuned rods, in the shape of a parabolic cylinder.[6] This success was commercially exploited when the decametric-wave telegraph service of the British Post Office was inaugurated in 1926, using curtain arrays beamed on Canada and South Africa.[7] Subsequent developments in this field are described in Section 4.

Between the two World Wars, steady progress was made in both the theory and the practice of aerial design. Shorter wavelengths were used as the longer wavebands became either overcrowded or unsuitable for the application in view; despite the attendant difficulties, higher-power transmitters became available as the result of improvements in valve and circuit technology. The related phase of aerial developments is described in Section 5.

An important landmark during the Second World War was the development of high-power generators of centimetric waves, and the associated radar systems. The required shaping of the beam of radiation was achieved, not by individually feeding discrete radiating elements, but by techniques akin to those used in optics. Some of the resulting aerials (described in Section 6), e.g. mirrors, horns and lenses, are similar in principle to those used by Hertz and his followers, but they employ immensely improved techniques and are correspondingly superior in performance.

(3) MYRIAMETRIC-, KILOMETRIC- AND HECTOMETRIC-WAVE AERIALS

The myriametric and kilometric wavebands are used for long-distance communication, with the earth and ionosphere behaving somewhat like a waveguide. The kilometric and hectometric wavebands are used for short-distance communication, such as local broadcasting. These services involve transmission along, or nearly along, the ground; as the ground is virtually perfectly conducting at these wavelengths, vertical aerials are used to achieve the maximum field strength in the required direction for a given power.

(3.1) Idealized Vertical Aerial

There is as yet no explicit solution to what at first sight appears a simple problem: the calculation of the radiation from a vertical cylindrical aerial energized at the base. Unfortunately the current distribution along the aerial cannot be calculated exactly, so that it is necessary to idealize the problem to make progress. The earliest calculations[4] assumed the current distribution to be the sinusoidal standing-wave pattern we should get on an open-circuited transmission line of the same length, along which a wave is propagated with the velocity of light. Although the aerial radiates, whereas the transmission line does not, the assumed and actual current distributions do not differ appreciably, provided that the transverse dimensions of the aerial are small compared with its length. Another assumption made in calculating the radiation pattern was that the ground behaves as a flat perfectly conducting plane; its effect is then equivalent to that of an image of the aerial in the ground.

The radiation pattern of a simple vertical aerial does not change appreciably with height unless the height is at least $0{\cdot}5\lambda$. For kilometric and longer wavelengths the aerial height is usually only a small fraction of the wavelength. In these cases the only point of using a high aerial is to increase the radiation resistance, thereby reducing losses in the ground and the coupling network (see Section 3.2.1). The radiation resistance can also be increased by adding a horizontal wire, or wires, at the top of the aerial (usually called a capacitance top).

In the hectometric waveband it is practicable to reduce upward radiation towards the ionosphere by making the aerial between $0{\cdot}5$ and $0{\cdot}6\lambda$ high, and this is now the common practice for broadcasting stations required to have anti-fading characteristics. (The reduction in upward radiation necessarily means that the ground-wave field strength for a given power is increased, but an increase in height is not generally worth while to achieve this result alone.) However, hectometric-wave anti-fading aerials are expensive, and with a view to reducing the required height and at the same time providing a control of the vertical radiation pattern, two methods are in use. One employs a capacitance

top, and the other divides the mast into two insulated sections, with an inductance connected across the gap (series loading). In practice these techniques permit a reduction in mast height of up to 20%.

In the kilometric waveband an aerial having anti-fading characteristics would have to be approximately 900 m high, and the cost would be prohibitive. A more economical method is to use a number of low aerials in the form of a ring. If these are energized with in-phase currents, an additional central aerial is necessary to achieve the required shape of the vertical radiation pattern.[8] The ring aerials can also be driven in progressive phase,[9] when the aerials carry equal currents but the phase changes progressively round the ring by an integral multiple of 2π radians. Page has described the general properties of the in-phase and progressive-phase ring aerials.[10] Hitherto ring-aerial systems have not come into use because of their high cost, but one is now under construction in Sweden.[11]

'Shunt-feeding' has been proposed as a means of avoiding the need for a base insulator; the mast is earthed at the base, and energized through a wire attached at a higher point. The saving in cost is in fact small, and the radiation pattern and input impedance may be adversely affected by the feeding system, so that this is rarely used.

Two methods are used for calculating the power radiated by an aerial. One involves integrating the power radiated through a sphere whose radius is large compared with both the wavelength and the height of the aerial. This method was applied to the vertical aerial by Pierce.[4] Brillouin's induced-e.m.f. method[12] is equivalent to calculating the flow of power through a cylinder just enclosing the aerial. It is more powerful than the first method, as it also enables the mutual reactance between coupled aerials to be calculated.

(3.2) Practical Vertical Aerial

In practice departures from the idealized conditions assumed in Section 3.1 are associated with

(*a*) The imperfectly-conducting ground.
(*b*) The aerial current distribution.
(*c*) Site irregularities, i.e. departures from flatness.

(3.2.1) Effect of Imperfectly Conducting Flat Ground.

Imperfectly conducting ground results in a loss of power in the ground, particularly in the immediate vicinity of the aerial. In the early part of the century the loss was reduced by means of a 'counterpoise', i.e. a grid of wires erected just above the ground; for convenience, extensive buried earth systems are now used instead. Even with these, if the aerial height is only a small fraction of the wavelength, the ground loss may be substantial. For instance, the 775 ft high aerial working on 16 kc/s at Rugby has an earth system extending over about 0·1 mile[2], and yet 80% of the power is dissipated in the ground near the aerial. For aerials at least $\lambda/4$ high the loss in the ground near the aerial is negligible, provided that an extensive earth system is used. However, imperfectly conducting ground attenuates the distant ground-wave field considerably.

Imperfectly conducting ground also affects the vertical radiation pattern, but this is important only with anti-fading aerials and when necessary can be compensated for by feeding the aerial appropriately (see Section 3.2.2). Monteath[13] has shown that, in practical cases, the earth system has a negligible effect on the radiation pattern.

(3.2.2) Aerial Current Distribution.

Theoretical solutions for the aerial current distribution have been obtained for cases where the shape is chosen to be convenient for mathematical analysis, e.g. the sphere,[14] the prolate spheroid[14] and the biconical aerial.[15] On the other hand, the solutions applying to cylindrical aerials introduce approximations at some stage. The fundamental difficulty is that of satisfying the condition that the tangential component of electric intensity must be zero at the surface of the aerial. The most suitable mathematical form of the solution depends on the precise shape of the surface. It will be different, for instance, if the corners are slightly rounded rather than sharp, although we can see intuitively that this cannot affect the current distribution significantly.

It is not therefore surprising that there is no exact solution for the current distribution on a cylindrical aerial. Approximate solutions have been obtained by Hallén,[16, 17] Schelkunoff[15] and Böhm.[18] Bearing in mind the fundamental difficulty of the problem, the results obtained by the three methods are in reasonably good agreement. They may be summarized by saying that the current distribution approximates to a sinusoid with a reduced velocity of propagation, together with an additional component which blurs the zero in the sinusoidal distribution. These deviations from the distribution assumed for the idealized aerial (which increase as the radius of the aerial is increased relative to its height) are important only in so far as they affect the vertical radiation pattern of anti-fading aerials. The adverse effects may be overcome by a slight reduction in the height, in addition to energizing the aerial at the point where the current is maximum rather than at the base.[19] The additional effect of imperfectly conducting ground can be overcome by energizing the aerial at both this point and the base simultaneously.[19]

For an aerial of given height, the larger we make the lateral dimensions the greater is the frequency range for a satisfactory impedance, and the smaller the voltage at the base. These reasons have contributed to the gradual displacement of wire transmitting aerials by insulated mast radiators. The same benefits can be achieved by using instead a cage of wires.

A transmitting aerial for an anti-fading broadcasting service is shown in Fig. 1. The mast radiator has a capacitance top and

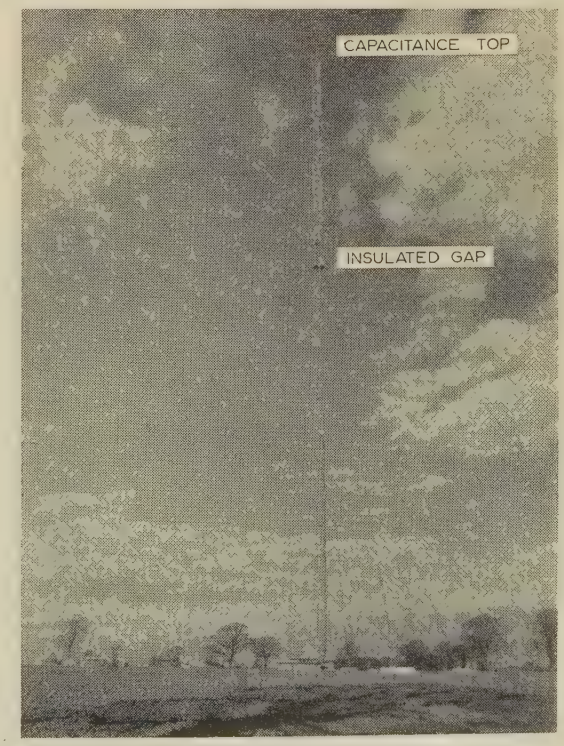


Fig. 1.—Anti-fading mast radiator.
Height of mast: 750 ft.
Operating wavelength: 464 m.

is fed with power both at the base and (by means of a transmission line running within the lower section) across the insulated gap.

(3.2.3) **Effect of Site Irregularities.**

So far, aerials radiating over flat uniform sites only have been considered. The effects of non-uniformity of the ground conductivity in different directions, and of the spherical shape of the earth, are generally negligible, except in so far as they affect the ground-wave field strength. On the other hand, irregular ground in the neighbourhood of the aerial modifies the vertical radiation pattern; an irregularity scatters the incident radiation and, if sufficiently large, imposes a ripple on the vertical radiation pattern. Page and Monteath[19] have described measurements made on small-scale model aerials radiating over plateaux and conical hills; they consider that, if a close approximation to the radiation pattern of an idealized anti-fading aerial is required, the undulations of the ground should not exceed $0{\cdot}05\lambda$ up to a range of at least 5λ from the transmitting aerial.

(3.3) Coupled Aerials

It is sometimes necessary to use a directional horizontal radiation pattern in order either to increase the field strength in particular directions without increasing the transmitter power or to avoid interfering with other users. Directivity is achieved by using two or more vertical radiators, the relative amplitudes and phases of the currents flowing in them being chosen to give the required pattern; some of the radiators may be parasitic, i.e. energized only by radiation coupling. When the current relationships have been decided, it is necessary to be able to calculate the effective impedance of individual radiators in order to design the coupling networks. The degree of coupling is expressed in terms of the mutual impedance between the constituent aerials; it can be calculated by an extension of the induced e.m.f. method.[12] The result for two parallel aerials has been given by Brown.[20]

(4) DECAMETRIC-WAVE AERIALS

The decametric waveband is used mainly for communication systems depending on reflection of the propagated signal from the ionosphere. The lower frequencies are used for ranges up to about 4000 km, which involve a single reflection (or hop), and the higher frequencies are used for multi-hop circuits and give world-wide communication. Because of the high attenuation over such long paths and the consequent degradation of the service by noise and other forms of interference, it is desirable to concentrate the power into a narrow beam which 'illuminates' only the area to be served. The radiation must be concentrated not only in the horizontal plane but also in the vertical, over a range of angles corresponding to the most favourable mode of ionospheric propagation. For a broadcasting service the angular spread of the beam is determined by the size and location of the area to be served, e.g. a nearby country may require a relatively broad beam projected at a large angle to the horizontal; a distant country generally requires a relatively narrow beam projected at a small angle to the horizontal. For a point-to-point service, on the other hand, the maximum spread of the beam is determined only by the uncertainty regarding the ionospheric path. The ionosphere sometimes causes a lateral deviation of the beam; variations in the height of the reflecting layer, and the diffuse nature of the reflection, also make it difficult to calculate the optimum projection angle precisely. Moreover, it is not always practicable to use a beam as sharp as is theoretically desirable if this demands an unduly expensive aerial. Consequently the same types of aerial are often used both for broadcasting and for point-to-point circuits.

The directional characteristics of the aerial are achieved by using assemblies of radiating elements, energized in such a way that the contributions from individual elements reinforce each other in the required directions. The success achieved in concentrating the radiation is defined by the gain, i.e. the ratio of the power supplied to a reference aerial to that supplied to the array, both giving the same maximum field strength. It is theoretically possible to obtain unlimited gain from arrays having overall dimensions as small as we please (see Section 7.4), but the radiation resistance of such arrangements is so low as to make them quite impracticable. The currents in the elements would be excessive and the losses consequently prohibitive, so that the theoretical gain cannot be realized. In practice, therefore, the higher the gain that is required the larger the radiating aperture must be, but there are many ways of arranging radiating elements to give a required pattern. In some cases it may be necessary to compromise between the complexity of the aerial system and its performance. This may involve considerations not only of technical performance, but also of mechanical design, the area of land required, the cost and other factors. There is therefore no 'best' aerial for any application; the advantages and disadvantages of possible ways of meeting the requirements must be studied in relation to the particular case.

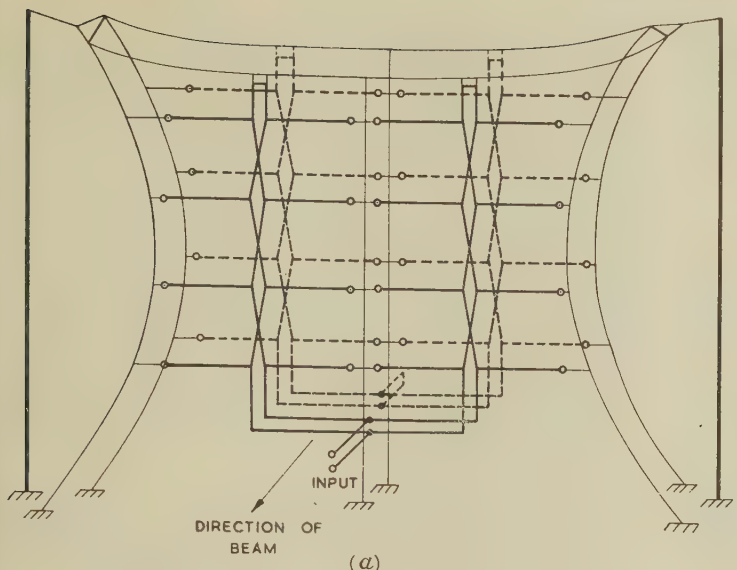
The first arrays used commercially were designed by Franklin and comprised a curtain of vertical radiating elements and a similar parasitic reflector curtain.[7] Arrays of horizontal elements are generally used nowadays;[21] this is because the ground behaves virtually as if perfectly conducting for horizontally polarized waves, whereas for vertically polarized waves there may be considerable loss of power in the ground.

The ground plays an important part in determining the elevation of the main beam, and it is usual when designing an array to assume that the site is flat. Irregularities of the terrain, in illuminated by the array, affect the radiation pattern as described in Section 3.2.3. If the ground has a uniform slope the beam is tilted to the same extent.

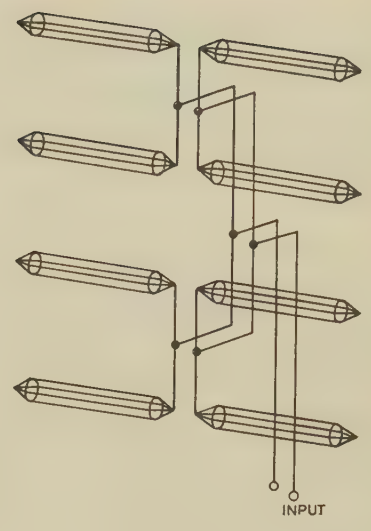
(4.1) Curtain Array

A typical curtain array is shown in schematic form in Fig. 2(*a*). Horizontal radiating elements are arranged in the form of a curtain, suspended from guys between the supporting masts. The elements carry equal cophased currents so that their contributions reinforce each other in the two directions normal to the curtain. This arrangement is sometimes referred to as a 'broadside' array. By adding a similar reflecting curtain the array can be made to produce a single main beam, together with a number of smaller side lobes. The reflector curtain is spaced $\lambda/4$ behind the driven curtain; it is usually energized only by radiation coupling and is tuned to minimize backward radiation. The array shown consists of four tiers spaced $\lambda/2$ apart, each tier comprising four collinear $\lambda/2$ elements. The transmission line is transposed between tiers so that the voltages applied to each pair of radiating elements are of the same amplitude and phase. By increasing the width of the array the horizontal beamwidth can be reduced; the number of tiers and mean height of the array control the vertical beamwidth and the projection angle. The ability to control these parameters independently makes the curtain particularly suitable for broadcasting, and it is widely used for this purpose.

By interchanging the functions of the two curtains the direction of the beam can be reversed. It is also possible to slew the beam over a small range of horizontal angles by displacing the input connection from the central point, since the currents in one half are then advanced in phase relative to those in the other. Slewing is achieved only at the expense of an increase in the side-lobe radiation, which may cause interference with other



(*a*)



(*b*)

Fig. 2.—Curtain array.
(*a*) Series feeding.
(*b*) Branch feeding.

services. The increase in side-lobe radiation with slew angle can be minimized by phasing the individual columns of radiators progressively, rather than in two groups. If, in addition, the amplitudes of the currents in individual columns are appropriately graded, an even greater reduction in side-lobe radiation can be achieved (see Section 5.4).

Curtain arrays of the type described are satisfactory over a frequency band of only a few per cent; beyond this range the gain falls rapidly owing to the change in the current distribution between elements. In addition, mismatching reduces the power-handling capacity. By using the branch-feeding arrangement shown in Fig. 2(*b*) these difficulties are largely overcome; a substantially symmetrical current distribution is achieved, and the matching is improved by using cages for the radiating elements. The bandwidth of arrays of this form is limited only by the change of the radiation pattern with frequency, or by power-handling capacity; they can be designed to cover a frequency range of about 1·25 : 1.

In an adaptation of the curtain array the elements are energized in the way already described, but the curtain is arranged parallel to, and about $\lambda/4$ above, the ground. The radiation from the curtain and its image reinforce each other in the vertical direction and produce a beam directed vertically upwards. Provided that an appropriate wavelength is used, the radiation is reflected at the ionosphere; these 'vertical-incidence' arrays are therefore used for broadcasting in regions where wide local coverage is required but a relatively cheap system must be employed.

(4.2) End-Fire Array

The radiating elements in an end-fire array (Fig. 3) are arranged in a line parallel to the ground, and spaced at equal intervals of not more than $\lambda/4$. The currents are generally made equal and in progressive phase. The contributions from individual elements at a great distance from the array are thus in phase in the direction in which the elements are spaced. The radiation pattern then takes the form of a main lobe in this direction, together with a number of side lobes. The angular spread of the main beam depends on the length, and the projection angle on both the length and the height, of the array. An advantage of the end-fire arrangement is that the height of the masts required to give a specified projection angle is less than for a curtain array. On the other hand, the beamwidths in the horizontal and vertical planes are interdependent, and this may sometimes be a disadvantage.

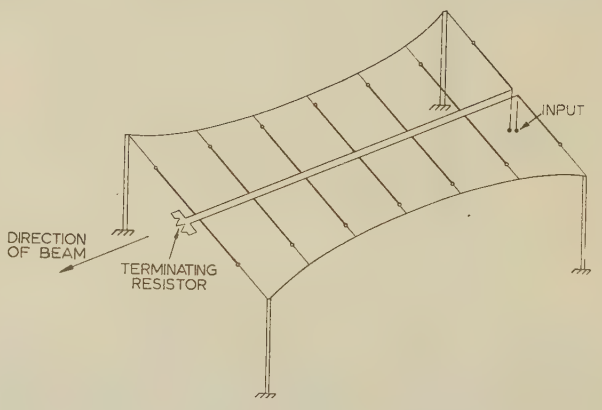


Fig. 3.—End-fire array.

The progressive-phasing requirement is usually achieved by terminating a transmission line in its characteristic impedance and then coupling the radiating elements very weakly, so as not to affect the matching of the line appreciably. This is a convenient but necessarily very inefficient arrangement, since an appreciable fraction of the power is dissipated in the terminating resistor. In one arrangement (the 'fishbone'), the weak coupling is achieved by connecting each radiating element through a small capacitance. In another (the H.A.D., or horizontal array of dipoles), very thin centre-fed dipoles λ in length, which have a high driving-point impedance, are connected directly across the transmission line. These forms of end-fire array are sometimes used for reception in those cases where it is more important to discriminate against interference than to supply the maximum voltage to the receiver. As in all aerials including a resistive element, the gain depends not only on the radiation pattern but also on the power dissipated in the resistor.

It is appropriate to mention here that the Yagi aerial,[22] which is commonly used for reception in the metric waveband, approximates to an end-fire array. Only one element is directly ener-

gized; the others are energized by radiation coupling only, the length of each being such as to give the best compromise between the amplitude and phase of the induced current. There is usually one parasitic reflector (i.e. a rod behind the driven element), and up to 16 parasitic directors (i.e. rods in front of the driven element), the gain increasing with the total number of rods used.

(4.3) Travelling-Wave Aerial

The travelling-wave aerial, or 'wave antenna', invented by Beverage[23] was a horizontal wire about λ in length, a few feet above ground level, erected in the direction of the wanted station; it was terminated in a suitable resistance to produce a travelling wave of current along the horizontal wire. It was originally proposed for reception of long-wave signals, and depended for its operation on the forward tilt of the incident wavefront. It is now mainly of interest in relation to the rhombic aerial, which may be regarded as a combination of four travelling-wave wire aerials. Each is equivalent to a continuous arrangement of collinear doublets in end-fire formation, so that the radiation is concentrated along a conical surface having the wire as axis.

(4.4) Rhombic Aerial

A rhombic aerial (Fig. 4) may be regarded as a transmission line which has been 'opened out' to allow it to radiate. When

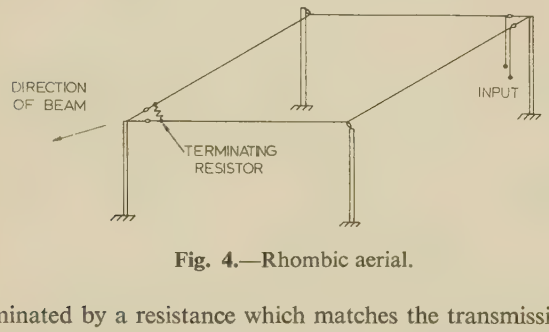


Fig. 4.—Rhombic aerial.

terminated by a resistance which matches the transmission line, the current distribution in each side of the rhombus approximates to a travelling wave. The angle of the rhombus is chosen to make the contributions from individual sides reinforce each other in the direction of the major axis. An expression for the theoretical radiation pattern has been derived by Foster.[24]

The rhombic aerial is widely used for both transmission and reception in point-to-point services. It is very suitable for this purpose, since it is mechanically simple and can be designed to have both a high gain and a relatively constant input impedance over a frequency band of about one octave. The changes in gain and beamwidth with frequency in typical cases have been given by Booth and MacLarty.[25] Charts to assist in such calculations have also been published.[26]

The rhombic aerial is sometimes used for broadcasting in preference to a curtain array, because of its low cost and its ability to cover a wide frequency band. However, it has the disadvantage that power is disspated in the terminating resistor, so that the gain is less than for the curtain array having a similar radiation pattern. The loss in the termination can be reduced to about 30% of the input power by constructing the sides of the rhombus from two or three spaced wires in parallel. The characteristic impedance is thereby reduced and the current for a given input power increased.

Other disadvantages, compared with the curtain array, are that the widths of the main beam in the horizontal and vertical planes are interdependent and change with frequency. In addition, the beam cannot be slewed (although it can be reversed by interchanging the input and the termination). The radiation pattern of the rhombic aerial is characterized by a large number of side lobes of appreciable amplitude. This means that, when transmitting, the aerial is more likely to cause interference with other services working on the same frequency, or be subject to interference when used for reception, than if these side lobes were absent.

The radiation pattern of the rhombic aerial is usually calculated for the idealized case in which the current along the wires is assumed to be constant, whereas in practice the current decreases towards the termination as the result of radiation. The practical performance does not in fact differ significantly from that of the idealized aerial, except in respect of the fine structure of the radiation pattern. However, there is a limit to the length of the rhombus that it is worth while using; in practice a side length of about 8λ is rarely exceeded. The gain of such a rhombus over an isotropic source (i.e. one radiating uniformly in all directions) is about 23 dB.

From time to time proposals have been made to improve the radiation characteristics of the rhombic aerial, e.g. by using them in broadside or end-fire formation (either with separate feeds to each rhombus or alternatively with one rhombus terminating the adjacent one). These more complicated arrangements are not used to any great extent in practice.

(4.5) Steerable Array

The utility of an array of fixed elements is greatly increased if it is possible to slew the beam of radiation by electrical means. The slewing and reversing of curtain arrays described in Section 4.1 is a simple example of this technique, but more ambitious systems are practicable for arrays intended only for reception.

One arrangement, the multiple-unit steerable antenna (m.u.s.a.),[27] which has been in use for the past 20 years on the important transatlantic service, minimizes fading by adjusting the response of the array to correspond to the optimum downcoming angle of the incident radiation. There are 16 horizontal rhombi in end-fire formation, the total length being about 2 miles. The outputs of individual rhombi are connected through adjustable phase-delay networks and then combined; the setting of the relative phase delays determines the angle of maximum response. The outputs from individual rhombi are connected to two separate groups of phase-delay networks; one group continuously sweeps over the whole range of phase delay and indicates the optimum projection angle, and the other is then set manually to correspond to this angle.

A steerable array which is more flexible than the m.u.s.a. uses a large number of vertical aerials, distributed over a circular area.[28] Each vertical aerial is connected to a wideband amplifier provided with a number of parallel output channels. The corresponding channels from all aerials are added, the reative phases being adjusted to give maximum response from the required direction. Signals may thus be received simultaneously from different directions and on different frequencies.

(5) METRIC- AND DECIMETRIC-WAVE AERIALS

The metric and decimetric wavebands are generally used for short-distance links involving ranges up to the 'radio horizon', i.e. a little beyond the optical range. These applications include point-to-point communication, television and sound broadcasting, radio navigation and telemetering. Some long-distance links using scatter transmission also operate in these wavebands.

The design methods applying in the longer wavebands are of course still applicable; assemblies of elements can be arranged to give the required radiation characteristics, and their impedance can be calculated. However, in the higher part of the decimetric

waveband the constituent $\lambda/2$ elements become so small that methods similar to those used in optics are more appropriate for shaping the radiation pattern, this change in technique occurring at about 1 000 Mc/s. It is therefore convenient to consider in this Section only those aerials involving assemblies of discrete radiating elements; the aerials involving techniques akin to those used in optics are dealt with in Section 6.

The aerials at both ends of the link will usually be at least λ above ground level. Furthermore, in many practical problems the permissible height of the receiving aerial is small, so that we are usually concerned with propagation along, or nearly along, the ground. In this case, to achieve the maximum field strength with a given transmitter power, the transmitting aerial must be as high as practicable. This means that both a high transmitting site and a high mast should be used, the mast height being determined by economic rather than technical considerations. Regulations for the safety of aircraft may sometimes impose an upper limit.

We can look at this in another way. Since the reflection coefficient of the ground at grazing incidence is -1, the radiated beam is elevated at a small angle. By increasing the height of the transmitting aerial we can depress the beam and so increase the received field. However, we always have the disadvantage of working on the 'skirt' of the vertical radiation pattern.

For some applications, the aerial must be a relatively simple mechanical structure such as a $\lambda/2$ dipole, an H-aerial consisting of a $\lambda/2$ dipole and reflector, or a Yagi aerial comprising a $\lambda/2$ dipole, reflector and one or more directors. Where a greater degree of mechanical complexity is permissible (whether for reception or transmission) highly directional arrays can be designed which are suitable for mounting on a single mast. A common arrangement is a number of vertically spaced tiers, each incorporating $\lambda/2$ elements, the tiers being energized in phase through coaxial lines in branch formation; this is similar in principle to the curtain array described in Section 4.1. The directivity in the vertical plane is controlled by the number of tiers, and in the horizontal plane by the number and disposition of the elements in each tier. If uniform radiation in all horizontal directions is required, the problem is to obtain a reasonable approximation to this condition, taking into account the effect of the supporting mast. For a directional horizontal radiation pattern the supporting mast may be used as an aperiodic reflector.

(5.1) Radiating Element

(5.1.1) The $\lambda/2$ Dipole.

The usual radiating element is a horizontal or vertical $\lambda/2$ dipole, because it is easy to support and the impedance is convenient for matching to the feeder system. In addition, it is often important to design the array to have a constant impedance over the working band of frequencies to avoid impairing the waveform of the radiated signal. This means that each constituent element must also have a relatively large bandwidth. One way of achieving this result is to increase the lateral dimensions of the element. We can also 'fold' the element, energizing only one branch (Fig. 5). This is equivalent to an impedance transformation and the addition of a compensating susceptance at the driving point.[29] Susceptance compensation without an impedance transformation can also be achieved by adding, at the central driving point, a transmission-line stub, which may also be used as a mechanical support.

It is usually convenient to energize the radiating elements through a coaxial rather than a balanced transmission line. For this purpose various types of 'balun' are available for changing from a balanced to an unbalanced impedance. They may also be designed to transform the impedance, or to increase the impedance bandwidth, or both.

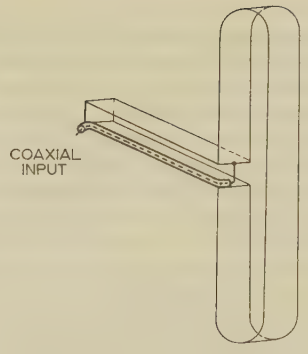


Fig. 5.—Wideband folded dipole.

(5.1.2) The Helix.

A wire many wavelengths long in the form of a helix [Fig. 6(*a*)] is a convenient radiating element, as it can have considerable directivity while requiring only a single feed-point. As a result of radiation, the current flowing along the wire is attenuated, so that the distribution approximates to a damped progressive wave. The radiating properties depend on the geometry of the helix.[30] In the axial mode of radiation the contributions from individual turns reinforce each other along the axis, so that the radiation is beamed in this direction and is circularly polarized. The axial mode is maintained over a relatively wide frequency band. On the other hand, if the turn-length is an integral number of wavelengths there is appreciable radiation normal to the axis, which is uniform in all directions and linearly polarized. If two collinear helices of opposite rotation are used, and fed as shown in Fig. 6(*b*), radiation in directions other than normal to the axis is substantially reduced. This arrangement is sometimes used in the U.S.A. for decimetric-wave broadcasting. A number of such elements can be mounted one above the other, and energized from a transmission line within the central support pole.

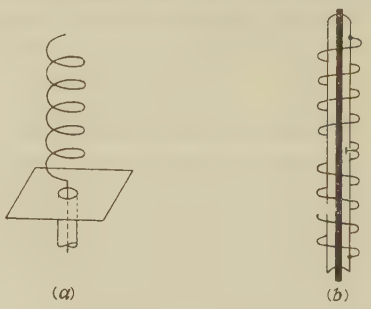


Fig. 6.—Helical aerial.

(*a*) Single helix.
(*b*) Two helices of opposite rotation.

(5.1.3) The Slot.

A slot cut in a metallic sheet acts as a radiating element if a generator is connected across the gap; however, because of its mechanical form the slot is used mainly at metric and shorter wavelengths. This radiator is best understood by reference to the concept of complementary aerials (analogous to Babinet's principle in optics). Corresponding to any reference radiating system there exists another in which the electric force is propor-

tional to the magnetic force in the reference system, and vice versa, the two systems being said to be complementary. An electric dipole consisting of a cylinder energized across a gap, and a magnetic dipole consisting of a cylinder of magnetic material energized by a loop encircling the material at the corresponding point, are complementary aerials. The ferrite rod used in domestic receivers is an example of a magnetic dipole. Circular and rectangular slots are used as radiating elements, but the rectangular form is more common and is the only one considered here.

Booker[31] showed that, if we have an electric dipole in the form of a thin strip of metal energized across a gap at the centre, the complementary aerial is a slot of the same dimensions cut in an infinite metallic sheet. Since a vertical dipole radiates vertically polarized waves, a vertical slot radiates horizontally polarized waves, and vice versa. Slots can be cut in sheets of finite size, although the size and shape of the sheet affect the radiation characteristics. A slot can also be boxed in on one side of the screen, thus confining the radiation largely to the other side, e.g. boxed slots are used in the skins of aeroplanes to provide dragless aerials. The radiation pattern, which depends on the shape of the surface in which the slot is cut, is generally determined by measurements on small-scale models. Theoretical results have been obtained for some idealized shapes.[32] Slots in cylinders are frequently used for metric- and decimetric-wave broadcasting (see Section 5.5) and slots in the walls of waveguides as the elements of centimetric-wave arrays (see Section 6.5).

(5.2) Effect of the Support Mast

If an aerial is mounted on a metallic mast some of the energy is reflected from the supporting structure, and the radiation pattern is consequently affected. A theoretical solution has been obtained for the re-radiation from an infinitely long cylinder.[33] In this case the analysis is simplified by considering the aerial as receiving; we suppose a transmitting doublet to be at a large distance from the cylinder and calculate the electric force it produces at a point near to the cylinder. This electric force is proportional to the voltage that would be induced in a receiving doublet located at the point in question; its variation with direction from the transmitting doublet gives the required radiation pattern.

The performance of an array is determined by adding the contributions of individual elements; arrangements giving the required shape of radiation pattern can then be devised.

An approximate method for a non-cylindrical mast, which is satisfactory if the cross-sectional dimensions are small compared with the wavelength, is to take as the equivalent cylinder one having the same capacitance per unit length. More accurate methods for dealing with larger masts of triangular and square cross-section have been published,[34] but they involve laborious computation.

(5.3) Multi-Element Aerials

The type of element from which the array is to be constructed having been decided, with the effect of the supporting structure taken into account, the problem is to arrange these elements over an aperture to give the desired radiation pattern. A frequent need is to achieve the maximum gain in a particular direction. The elements are then arranged in a plane normal to this direction and fed with in-phase currents, the width and height of the array determining the beam width in the two principal planes. There is a spacing between the elements which gives the maximum gain. For $\lambda/2$ dipoles, for instance, the optimum spacing is about $0{\cdot}7\lambda$ for parallel and λ for collinear elements, the precise value depending to some extent on the number of dipoles. (The $\lambda/2$ spacing used for the decametric-wave broadside arrays described in Section 4.1 is for convenience in feeding the tiers rather than to maximize the gain.)

The radiation pattern of a broadside array in one of the principal planes is the same as that for a line of similarly distributed sources—generally called a linear array. The pattern exhibits a main lobe directed normal to the line of elements, together with a number of side lobes and intermediate zeros; the longer the linear array, the narrower is the main lobe and the greater are the number of side lobes. For point-to-point working this would be satisfactory, but in some applications it may be necessary to control the overall shape of the radiation pattern. For instance, to increase the gain of the aerial in television broadcasting, a broadside arrangement of similar vertically spaced tiers is used. However, if the number of tiers is very large (up to 50 are used in the decimetric band) and they are fed with equal co-phased currents, there will be gaps in the coverage near the transmitting aerial on account of the zeros in the vertical radiation pattern. Ideally we should illuminate the service area uniformly and suppress all radiation in directions above the horizon. The requirement is similar for an airborne radar set, which must uniformly illuminate the earth's surface ahead. In these two cases, as there is a preferred shape of radiation pattern, there is also an upper limit to the gain of the aerial even though no limit is imposed on the number of tiers. A rather different example is that of a radar aerial required to scan the sky. If the radiation pattern had appreciable side lobes there would be ambiguous indications. It is therefore necessary in this case to minimize side-lobe radiation, even though this entails a slight widening of the main lobe and consequently reduced resolving power.

(5.4) Shaping the Radiation Pattern of a Linear Array

The radiation pattern of a linear array and the feed distribution (i.e. the currents supplied to the radiating elements expressed as a function of the distance along the array) are related by a pair of Fourier transforms. The radiation pattern can be derived directly from the feed distribution.

It is sometimes necessary to carry out the reverse process, i.e. to determine the feed distribution from the radiation pattern. In so doing it is necessary to specify the phase as well as the amplitude of the radiation pattern, but since the phase is generally unimportant there is not a unique solution for the feed distribution. To make progress it is therefore usual to specify arbitrarily that the phase of the radiation at a given distance is independent of direction. The calculated feed distribution in general extends to infinity whereas in practice the permissible length of the array is limited. The best we can do, therefore, is to approximate to the requirement, the greater the radiating length the better being the degree of approximation achieved. Consider, for example, the radiation pattern for a uniform co-phased distribution 10λ long, shown by the full line in Fig. 7. It consists of a main lobe together with a number of side lobes and is analogous to the optical diffraction pattern obtained by illuminating a slit. Such a pattern would be unsatisfactory for a radar aerial required to scan the sky; for this application a single 'pencil' beam with no side lobes is desirable. The Fourier transform of such a pattern shows that a tapered (rather than a uniform) amplitude distribution of infinite extent is required. If we decide for practical reasons to restrict the radiating length, small ripples are added to the radiation pattern, but these are nevertheless much smaller than those corresponding to the uniform distribution. Fig. 7 shows, for instance, the radiation pattern for a triangular amplitude distribution. It has appreciably lower side lobes than that for the uniform distribution. A distribution can in fact be found which has a radiation pattern

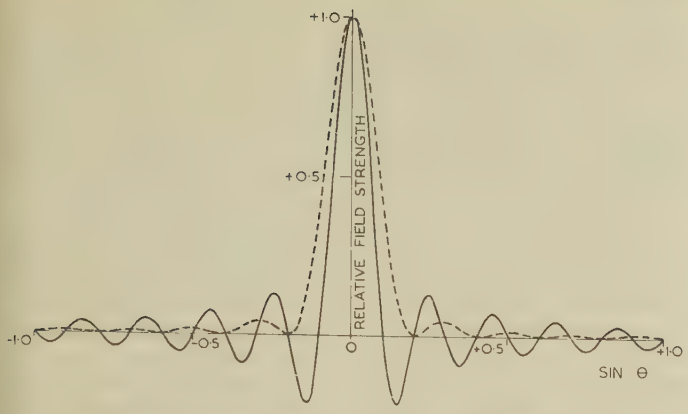


Fig. 7.—Radiation pattern of 10λ co-phased linear array.
—— Uniform amplitude distribution.
---- Triangular amplitude distribution.
θ = Angle to normal.

with no side lobes at all, but the beam is then considerably wider and the gain correspondingly smaller. Dolph[35] has shown how to calculate the distribution giving the narrowest beam width for a specified side-lobe level, when using equally spaced discrete sources at least $\lambda/2$ apart, carrying co-phased currents. (If the sources are spaced less than $\lambda/2$ the super-gain design principle described in Section 7.4 is applicable.)

The more general case of approximating to an arbitrary radiation pattern has been investigated by Woodward[36]; he showed that for a linear array of $n\lambda$ the pattern can be specified at $(2n + 1)$ points. An example is illustrated in Fig. 8. The

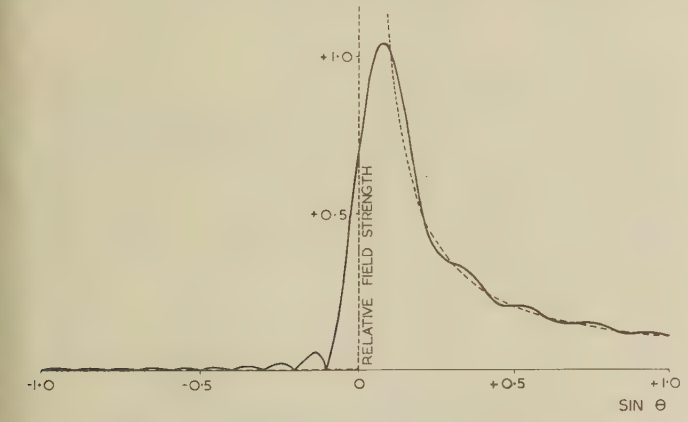


Fig. 8.—Radiation pattern of a 10λ linear array.
---- Ideal radiation pattern.
—— Actual radiation pattern.
θ = Angle to normal.

broken line shows the ideal radiation pattern for an airborne radar system required to illuminate uniformly the earth's surface ahead, and the full-line curve is the approximation obtained using a 10λ linear array. Woodward's method makes the arbitrary assumption that the phase of the distant field is the same in all directions. Ksienki's method[37] is not so restricted but involves a certain amount of trial and error.

The above discussion has been mainly in terms of a continuous feed distribution, but the same general considerations apply for uniformly spaced discrete sources provided that the spacing is not appreciably greater than $\lambda/2$.

(5.5) Aerials having a Uniform Horizontal Radiation Pattern

Some aerials, e.g. those for broadcast transmissions, are required to have a substantially uniform horizontal radiation pattern. We may regard a single tier of such an aerial as an 'element' for building up into a multi-tier array. A method of meeting the requirement is to use a continuous distribution of radiating elements (vertical, tangential or radial) in the form of a ring round a cylindrical mast. This arrangement is similar to the ring aerial described in Section 3.1 but with the support mast constituting a central parasitic reflector. The elements carry equal currents which may be in phase, or in progressive phase, i.e. the phase changing uniformly round the ring, the total phase change being an integral multiple of 2π radians. From considerations of symmetry, the horizontal radiation pattern is uniform for either arrangement. A continuous distribution of radiating elements is impracticable, and a compromise must be sought between the number of elements and the uniformity of the radiation pattern.

The most suitable arrangement depends on the particular case. A progressive-phase system is often used if the vertical support mast has a relatively small cross section, e.g. a thin pole mounted on top of the main structure. For vertical polarization the elements are disposed parallel to the central pole. For horizontal polarization each tier may consist of two horizontal dipoles at right angles carrying equal currents in phase quadrature, which we can regard as four radial elements carrying currents in progressive phase. This type of aerial, shown in Fig. 9, is known

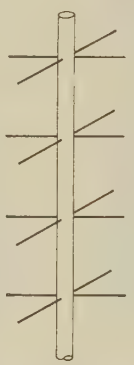


Fig. 9.—Four-tier turnstile aerial.

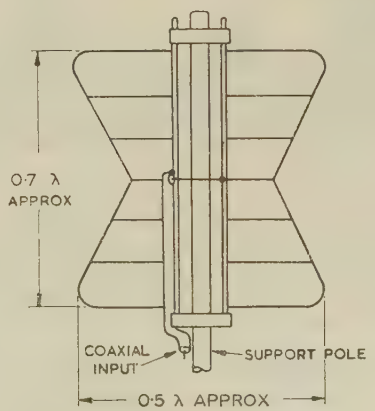


Fig. 10.—Batwing element.

as a 'turnstile'. In a modification of the turnstile, the radiating element consists of a skeletonized 'batwing' (Fig. 10); this can be regarded as a short-circuited transmission line (or slot) across which radiating elements are tapped, the lengths being graded to spread the current distribution. As a result, one tier (com-

prising two batwings at right angles) has approximately the same gain as two tiers of the turnstile aerial. By a suitable choice of the slot length and width and the shape of the batwing, the impedance can be made much less frequency-dependent than that of a simple dipole. The batwing aerial is commonly used in the metric waveband for horizontally polarized television transmissions.

For masts of large lateral dimensions an in-phase ring of vertical or horizontal dipoles is usual, the number required to achieve a uniform horizontal radiation pattern increasing with the lateral dimensions of the support mast.

Longitudinal slots cut in a cylindrical mast may be arranged to give a substantially uniform horizontal radiation pattern. Three examples of slot aerials used for broadcasting are shown in Fig.11.

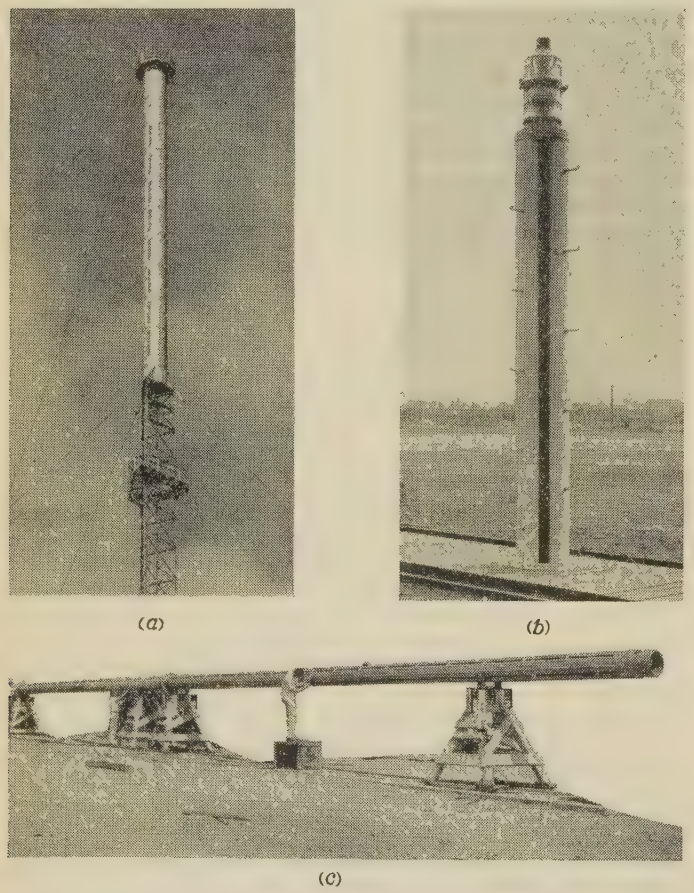


Fig. 11.—Slot aerials used for broadcasting.

(*a*) Slot aerial used for sound broadcasting in the United Kingdom (87·5–95 Mc/s). Each tier consists of four slots, equally spaced round a cylinder of 6·5 ft diameter.
(*b*) Slot aerial used for sound broadcasting in the United States (88–108 Mc/s). Each tier consists of single slot in a cylinder of 1·8 ft diameter.
(*c*) Slot aerial used for television broadcasting in the United States (470–890 Mc/s). Each tier consists of three slots, equally spaced round a cylinder of 1·3 ft diameter.

(6) CENTIMETRIC-WAVE (MICROWAVE) AERIALS

For the centimetric and shorter wavelengths, often collectively described as the microwave band, a radical change in the approach to aerial design becomes desirable, stemming mainly from the inconvenience of achieving directional characteristics by energizing a large number of separate radiating elements of small size. As the dimensions of the aerials are generally large compared with the wavelength, optical design principles are both preferable and practicable. A recent review of these techniques has been given by Harvey.[38] The aerials used are essentially collimating devices in which the approximately spherical wavefront of a primary radiator is transformed into a virtually plane wavefront by a secondary radiator. The two main classes are those which make use of the reflecting properties of metallic sheets and those which employ the refracting properties of materials, analogous to optical reflecting and refracting telescopes respectively. The metallic horn and dielectric aerial may be regarded as hybrids of these two classes. An array of slots is often used as the primary radiator in radar systems.

Microwave aerials are used for radio-relay and radar purposes, and the parameters of interest are the impedance, the gain, the beam width, and the ratio of the side lobes to the main lobe. It is necessary to decide first on the size and configuration of the aperture to give the required beam shape, and then to control the field distribution over this aperture appropriately, in both amplitude and phase. The radiation pattern may be calculated by using Huyghens's principle, in which the aperture distribution is regarded as equivalent to an infinite number of elementary sources spaced over the aperture in broadside formation. Booker and Clemmow[39] regard the aperture as radiating an angular spectrum of plane waves, each wave being associated with one of the Fourier components of the distribution of tangential electric force over the aperture. The radiation pattern of the complete spectrum can then be calculated. The calculation of the radiation from an aperture is the 2-dimensional equivalent of the calculation for the linear array, described in Section 5.4.

As with a linear array, the gain from a given aperture is a maximum if the distribution of the tangential component of the

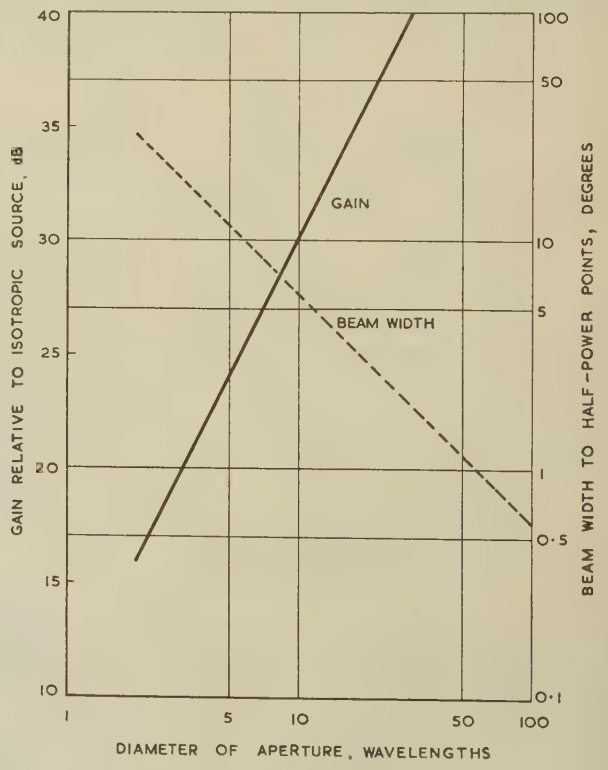


Fig. 12.—Gain and beam width of uniformly illuminated circular aperture.

field across it is uniform and cophased;* the larger the aperture the larger is the achievable gain. The gain and beam width of a circular aperture illuminated in this manner are shown in Fig. 12.

* This statement is strictly true only for distributions which vary relatively slowly over the aperture; in theory the gain can be increased indefinitely if the distribution is allowed to vary rapidly in both magnitude and sign, but this is generally impracticable (see Section 7.4).

as a function of the diameter. For a narrow-beam array the level of the largest side lobe is −17·6 dB relative to the principal maximum, independent of the diameter of the aperture. For some applications this level would be too high. It can be reduced by tapering the amplitude distribution towards the edge of the aperture (as described for linear arrays in Section 5.4) at the expense of a slight increase in the beam width and a corresponding reduction in the gain. Fig. 12 gives the upper limit for the gain of a uniformly illuminated aperture. In practice the gain of the aerial will be rather less because the primary radiator either will not illuminate the aperture uniformly or, if it does, will illuminate other regions as well.

(6.1) Reflecting Aerials

(6.1.1) Paraboloid of Revolution.

A metallic surface generated by the revolution of a parabola about its axis and illuminated by a source placed at the focus is similar in principle to the corresponding optical mirror; it is usually referred to as a paraboloid or 'dish'. Paraboloids have the advantage that they can be used over a wide range of frequencies, although generally the primary radiator requires changing to suit the operating frequency. For metric-wave paraboloids the combination of a $\lambda/2$ dipole and reflector is commonly used as the primary radiator; at shorter wavelengths an open-ended waveguide, a slotted waveguide or a metallic horn may be used. A helix operating in the axial mode is suitable if circular polarization is required.

A typical feed arrangement is shown in Fig. 13(*a*). One dis-

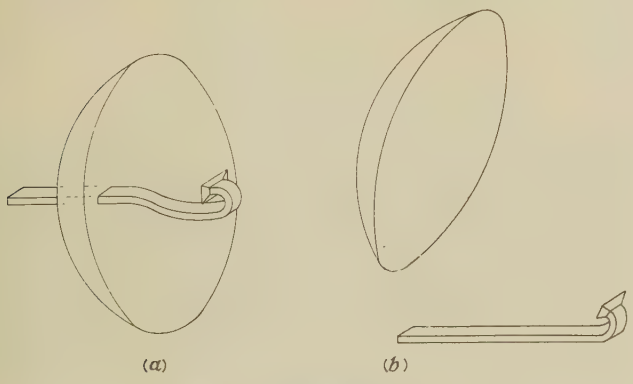


Fig. 13.—Paraboloid.
(*a*) Rear feed.
(*b*) Offset feed.

advantage is that energy may be reflected from the paraboloid back into the primary radiator, thus affecting the input impedance. This can be overcome by adding a 'vertex plate', i.e. a small plate mounted at the centre of the paraboloid and spaced $\lambda/4$ from it. The area of this plate is such that the signal it reflects into the primary radiator cancels that reflected from the remainder of the paraboloid. An additional disadvantage of the arrangement of Fig. 13(*a*) is that the waveguide feed to some extent obstructs the forward radiation. This can be avoided by using the offset feed, shown in Fig. 13(*b*).

All aerials including reflecting sheets, whether curved or flat, must be accurately shaped if the theoretical performance is to be achieved. The contour irregularities should not exceed one-sixteenth of the shortest wavelength used.

(6.1.2) Parabolic Cylinder.

The parabolic cylinder, Fig. 14(*a*), is convenient when it is required to fix the beam width in the two principal planes independently. One important practical example is the cheese

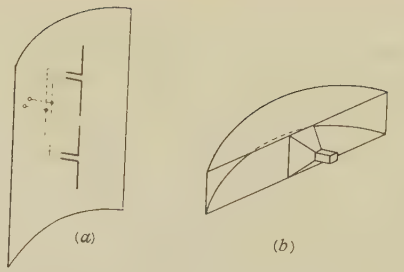


Fig. 14.—Parabolic cylinder.
(*a*) Linear-array feed.
(*b*) Cheese aerial with horn feed.

aerial, shown in Fig. 14(*b*). A slice (whose height is much smaller than its width) of the parabolic cylinder is enclosed between two parallel plates. The resultant fan-shaped beam is narrow in the horizontal plane and wide in the vertical. It is therefore particularly suitable for rotating radar systems where the principal requirement is good azimuthal resolution.

(6.1.3) Corner Aerial.

The corner, or V, aerial (Fig. 15) is a mechanically simple arrangement which radiates a beam directed normal to the axis.

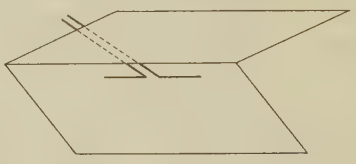


Fig. 15.—Corner aerial.

The beam width in the principal plane normal to the axis is determined mainly by the width of the sheets and the angle of the V; the beam width in the other principal plane is determined mainly by the length of the sheets and the type of primary radiator. The primary aerial may be, for example, a row of collinear electric dipoles, or a slotted cylinder. The only exact theoretical solution for the radiation pattern of the corner aerial is for sheets of infinite extent.[32] Wilson and Cottony have published gain and radiation-pattern measurements for a wide range of sizes of the reflecting sheets.[40, 41]

(6.1.4) Passive Reflector.

The term passive reflector is used to describe a metallic sheet not immediately associated with the aerial, e.g. a sheet mounted on a mast and illuminated by an aerial located at ground level. This arrangement avoids the necessity of providing an elevated aerial and a transmission line up the mast. Medhurst[42] has published the performance characteristics of flat sheets and Bedrosian[43] those of curved sheets. Curving the reflector produces a focusing effect so that the gain is greater than for a flat sheet of the same size.

A tiltable plane reflector is a convenient way of changing the direction of a beam from a more complicated fixed aerial system. One such reflector 260 ft long and 100 ft wide is used for radio-astronomy investigations.[44]

(6.2) Refracting Aerials—The Lens

In the lens, the wavefront from the primary radiator is shaped by interposing a medium having a refractive index different from unity. There are two main types. In the first, the velocity of propagation near the centre of the wavefront is decreased by interposing a medium having a refractive index greater than

unity. In the second, the phase velocity at the outer parts of the wavefront is increased by interposing a medium having a refractive index less than unity. The first type is analogous in operation and similar in form to the optical lens, whereas there is no optical analogy of the second.

A feature common to all lens aerials is that the primary radiator is behind the refracting medium. The feed to the primary radiator does not therefore obstruct the emergent wave as it does in the case of some reflecting aerials. The incident wave is, however, partially reflected at the refracting medium. Although reflection can be minimized by methods akin to those used in optics,[38] it is necessary to compromise by using refractive indices not very different from unity. Values used are about 1·5 for the first type and 0·6 for the second.

(6.2.1) **Media having a Refractive Index greater than Unity.**

A suitable material is any low-loss dielectric which is amenable to machining or moulding, such as polystyrene.* Another possibility is to change the permittivity from point to point within the lens, e.g. by using assemblies of polystyrene discs drilled with holes. All such lenses are bulky and heavy so that other arrangements are more attractive.

Kock[45] showed that similar results could be achieved using an 'artificial dielectric' in which small metallic particles, spaced a small fraction of the wavelength apart, are embedded in a low-loss material such as polystyrene foam. These media are lighter and therefore more convenient than homogeneous types. Their general properties have been described by Brown and Jackson.[46] We can regard the embedded particles as increasing the permittivity and so reducing the phase velocity of the medium. Kock's original proposal was to use spherical particles, but with this arrangement the maximum value of the refractive index which can be achieved is limited, even when the particles are densely packed. No such limit is imposed by discs arranged normal to the direction of propagation. If a material of large refractive index is used, however, the reflection loss at the incident surface will be prohibitive unless matching layers are used. One method is to provide an outer layer $\lambda/4$ thick, loaded with particles of reduced size.

Both the embedded sphere and disc dielectric have refractive indices which are independent of the polarization of the incident wave, but their characteristics change with frequency. Artificial dielectrics which are easier to manufacture can be used if the lens is required to work with waves polarized in one direction only, e.g. thin metallic strips can be used instead of discs.

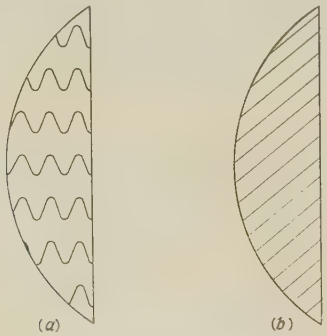


Fig. 16.—Path-length lens.
(*a*) Serpentine path.
(*b*) Inclined path.

Focusing of the incident radiation can also be achieved by means of the so-called 'path-length' media,[47] Fig. 16(*a*), the waves incident on the centre of the lens being made to follow a serpentine path. In practice it is more convenient to use metallic plates inclined at an angle to the direction of propagation as in Fig. 16(*b*). Generally, path-length media are more difficult to construct and do not give such a good performance as other types; they are consequently little used.

In a type of lens used for radio-link purposes in France,[48] the primary source illuminates a metallic sheet pierced with holes of diameter small compared with the wavelength, collimation being achieved by reducing the size of holes towards the edge of the aperture. (This aerial is complementary to one consisting of metallic discs of the same size and disposition as the holes in the metallic sheets.) In one lens, seven such sheets in cascade are used.

(6.2.2) **Media having a Refractive Index less than Unity.**

The parallel-plate lens consists of a series of parallel metal plates with edges in the direction of the incident electric force. The spacing of the plates is between one-half and one free-space wavelength, so that the phase velocity in the plate structure is greater than that in free space. By increasing the width of the plates towards the edge of the lens, as in Fig. 17, it is possible

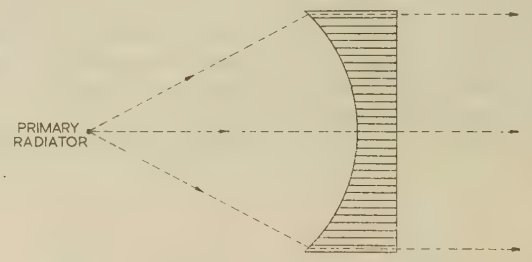


Fig. 17.—Cross-section of parallel-plate lens.

to transform the incident spherical wavefront into a plane wavefront. In practice it is not necessary continuously to increase the width of the plates towards the edge of the lens. The width of any plate can be reduced, provided that the combined free space and guided path is decreased by an integral number of wavelengths. This gives the 'zoned' shape shown in the spherical lens of Fig. 18, and considerably reduces the bulk and weight of the lens. Reflection at the incident and emergent surfaces may

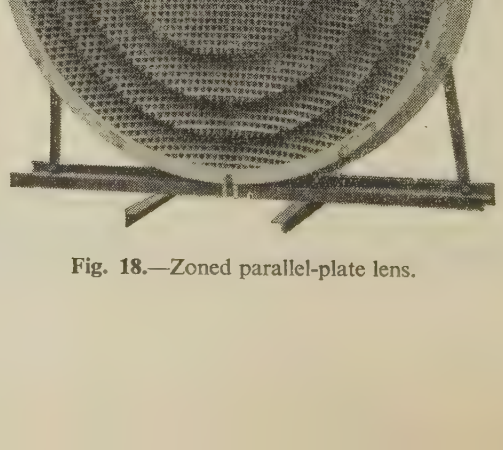


Fig. 18.—Zoned parallel-plate lens.

* Before 1900 Lodge had experimented with a similar type of lens made of pitch, and later one of glass. Righi tried paraffin and sulphur.

be minimized by additionally zoning alternate halves of the lens (or alternate plates) by $\lambda/4$. Plates perpendicular to the direction of the electric force do not affect the performance, so that the 'egg-box' type of construction shown in Fig. 18 may be used. If the lens is required to operate for waves of any polarization, the 'cells' should be of square cross-section.

An arrangement equivalent to the parallel-plate lens is a lattice of metallic rods in the same direction as the plates. It has the advantage that there is no need to make the spacing in the direction of the magnetic force greater than one-half the free-space wavelength, so that the possibility of high-order reflected waves is minimized. The rod radius becomes impracticably small for wavelengths less than about 10 cm.

The parallel-plate lens is satisfactory over only a limited band of frequencies, so that there is a tendency for its displacement by aerials capable of working over a wider band.

(6.3) The Horn

One of the simplest means of achieving a directional aerial is to flare out the end of the waveguide into a horn of the required aperture. Fig. 19(*a*) shows the sectoral horn, which is flared

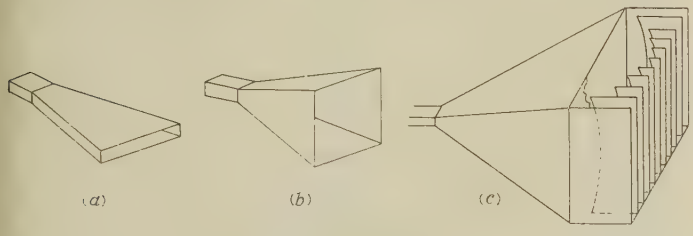


Fig. 19.—Metallic horn.
(*a*) Sectoral horn.
(*b*) Pyramidal horn.
(*c*) Phase-corrected horn.

out in one dimension only, and Fig. 19(*b*) the pyramidal horn. The radiation is guided by the walls; if the flare were sufficiently gradual the phase would be reasonably uniform over the radiating aperture, but the horn would then be inconveniently long. If the flare length is reduced the path from the throat to the centre of the radiating aperture becomes shorter than that to the edge, so that the phase front does not coincide with the aperture plane. This difficulty can be overcome by adding a series of parallel metallic plates in the direction of the electric force,[49] as shown in Fig. 19(*c*). These increase the phase velocity of the waves near the edge of the aperture, and so make it possible to achieve a virtually plane wavefront across the mouth. The phase-corrected horn may be regarded as a lens in association with a primary source whose radiation is guided by the walls of the horn.

In an alternative arrangement, shown in Fig. 20,[50, 51] the phase may be corrected by capping the horn with a sector of a paraboloid of revolution, with the apex of the horn at the focus. The advantage of this arrangement is that the same reflector can be used over a wide frequency band. It is akin to the paraboloid with offset feed, shown in Fig. 13(*b*).

(6.4) Dielectric Aerial

Dielectric aerials employ dielectric elements for guiding the waves, as distinct from conducting elements. An account of work in this field has been given by Kiely.[2] There are three main types of aerial, all of which may be excited by open-ended waveguides or probes:

(*a*) The dielectric rod [Fig. 21(*a*)], called the 'polyrod' since it is usually made from polystyrene.
(*b*) The dielectric tube [Fig. 21(*b*)].
(*c*) The dielectric horn [Fig. 21(*c*)].

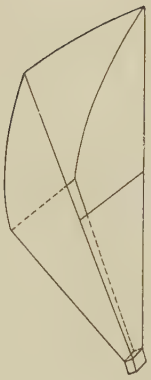


Fig. 20.—Horn reflector.

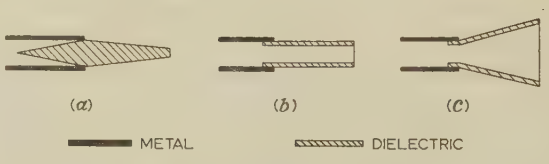


Fig. 21.—Dielectric aerials.
(*a*) Dielectric rod.
(*b*) Dielectric tube.
(*c*) Dielectric horn.

The dielectric rod is usually tapered, and measurements show that the radiation pattern is satisfactory over a wide band of frequencies and is relatively free from side lobes. Some workers regard the aerial as a leaky guide which radiates continuously along the length of the rod, and others as a guide for a wave directed along the surface of the rod so that it radiates mainly from an aperture plane normal to the axis of the rod; the theory cannot therefore yet be regarded as established. The dielectric rod has the advantage of sealing the aperture of the exciting system, but its performance is affected by surface moisture. It has been used in practical applications, but not extensively.

The theory of the dielectric tube involves consideration of the field discontinuities at both the inner and the outer surfaces and presents more difficulty than that of the dielectric rod. The radiation pattern takes the form of a single lobe for all thicknesses below a critical value, but is multi-lobed for greater thicknesses.

The dielectric horn resembles the metallic horn described in Section 6.3, with the metal walls beyond the throat replaced by dielectric walls. From the point of view of theoretical analysis it is an even more complicated structure than the dielectric tube, and little theoretical or experimental work on its performance has been published. It is more directive than the corresponding metallic horn of the same aperture.

(6.5) Array of Slots

Instead of coupling the waveguide to a separate aerial it is sometimes convenient to allow the guide itself to radiate. Slots are cut in one of the walls to interrupt the flow of current,[52] and the guide is terminated in a short-circuited section $\lambda/4$ in length. For instance, slots may be cut in the broad wall, as in Fig. 22(*a*). A common arrangement is to space the slots one-half of the guide wavelength apart and displace them either side of the centre-line, in order to reverse the polarity of the feed to adjacent slots. An alternative arrangement is to cut the slots in the narrow wall of the guide. The slots may be inclined in order to interrupt the current flow, as in Fig. 22(*b*), the change

Fig. 22.—Array of slots in rectangular waveguide.
(*a*) Slots cut in broad wall.
(*b*) Inclined slots cut in narrow wall.

of polarity of the feed to adjacent slots being achieved by reversing their inclinations. If the slots are not inclined, they must be separately coupled by means of probes.

These arrangements are similar to the broadside arrays described in Section 4.1 and suffer from the same disadvantage, namely that the performance is satisfactory over only a relatively narrow band of frequencies.

(7) SPECIAL TECHNIQUES

(7.1) Direction-Finding

The principles of direction-finding were being explored at the turn of the century. Brown[53] took out a patent in 1899 which depended on rotating a pair of oppositely coupled aerials (which gives a figure-of-eight radiation pattern) until a zero is obtained. The use of a rotating loop aerial for the same purpose was probably suggested by Lee de Forest. Later Bellini and Tosi used two crossed loops in association with a radiogoniometer.[54] The bearing ambiguity can be eliminated by associating with the loop a separate 'sensing' aerial, which can be arranged to convert the figure-of-eight reception diagram into a cardioid. The crossed-loop arrangement is still used in ships and aeroplanes although supplanted for accurate navigation by more modern aids. Ferrite-cored loops are also used. During the 1930s, crossed-loop systems were found to give errors in bearing at night, and this 'night effect' was traced to the horizontal electric-field component resulting from reflection at the ionosphere.[55] Attention was therefore turned to a system which had been proposed by Adcock in 1919[55]; it consists of two fixed pairs of oppositely coupled vertical aerials in association with a goniometer. The system has since been developed to give accurate results on much shorter wavelengths. A similar arrangement of spaced loops is also used, but mainly as a research tool. Such systems depend for their accuracy on the uniformity of the incident field over the receiving site, and this is adversely influenced by irregularities of the terrain. Extensive earth screens are sometimes erected just above the ground to overcome this difficulty. More distant reflecting objects still affect the bearing accuracy, but their effect can be partially taken into account by a calibration process.

The direction-finders so far discussed work on the principle of exploring the incident wave over a relatively small volume. Site errors can be reduced by using a wide-aperture system, in which the aerials in each pair of oppositely coupled elements are spaced a few wavelengths apart. These instruments can be made sufficiently accurate to use as research tools for investigating ionospheric irregularities. If an automatic plot of the bearing is required, one difficulty of the wide-aperture system is that there are ambiguities in the indication. These can be overcome by using a ring of aerials, signals proportional to the phase differences between adjacent aerials being fed in rotation to a receiver. The phase modulation relative to a reference signal gives the bearing information.[56]

A description of current direction-finding practice has been published by Hopkins and Pressey.[57]

(7.2) Radar

Radar systems are required to give information on the azimuth, elevation and distance of a target. The technique was developed originally for ionospheric sounding; it was later adapted for the location of aircraft and for navigational purposes. The distance is determined by measuring the delay of the reflected signal, using either a pulsed or a frequency-modulated transmission. The remaining problems are akin to those arising in direction-finding, but with the important difference that the radar installation must be able to work with a relatively weak reflected signal. The aerials therefore play a very important part in determining both the accuracy and the sensitivity of the system. Horizontal polarization is usual, because it gives rise to less ground clutter, i.e. permanent echoes from local obstacles. British radar practice up to 1946 has been described by Ratcliffe[58] and that of the United States up to 1947 by Friis and Lewis.[59]

There are two requirements peculiar to radar: one is the use of the same aerial for transmission and reception,[58] and the other is the necessity to scan the target area rapidly. Resolution in the horizontal plane is usually accomplished by rotating the complete aerial system. If resolution in the vertical plane also is required, the beam of the primary radiator must scan the secondary radiator, and this can be accomplished either by mechanical movement or by electrical beam slewing. For a reflecting-type aerial the usual method is to move mechanically the primary source in the focal plane, but this affects the beam width and also gives rise to side lobes. It is difficult, for instance, to swing the beam of a paraboloid through an angle greater than the beam width. The spherical mirror has a radiation pattern which is inferior to that of the paraboloid, but is less subject to aberrations provided that the source is moved along the correct arc; in this case, scans of many times the beam width can be achieved. Another method is to rotate the mirror about the feed as a centre.

With lenses, by a suitable choice of the shape of the surface, the thickness and the refractive index, relatively wide-angle scanning can be achieved (see Chapter 6 of Reference 60). A lens offering the possibility of unrestricted scanning but involving moving only the primary feed has been suggested by Luneberg.[61] Such a system must possess spherical symmetry, the refractive index varying with distance from the centre in such a way that the radiation from a source on the surface is transformed into an emergent plane wave. This is an attractive idea in principle, but its realization is complicated by the lack of suitable materials, so that the work so far done must be regarded as being in the experimental stage.[62, 63, 64]

(7.3) Radio Astronomy

In pursuance of the correspondence between electromagnetic and optical phenomena, Lodge, at some time between 1897 and 1900, endeavoured to detect r.f. waves emanating from the sun. Although unsuccessful, he was convinced that his failure was due only to instrumental limitations, but almost half a century was to elapse before his theory was substantiated. Radio stars are very weak sources of electromagnetic radiation, and the detecting instruments must be very sensitive and have a high resolving power. As a result, the aerials represent the major item of expenditure in radio-astronomical work.

A steerable paraboloid is commonly used for such investigations. It can be used over a wide band of frequencies (although this generally means changing the primary feed), and is thus a general-purpose instrument. One of the largest so far built is at the radio-astronomy station at Jodrell Bank, England; the diameter of the reflector is 250 ft, the focus being in the aperture plane. The size, and hence the resolving power, of such aerials is limited by practical considerations. For this reason, fixed aerial systems are frequently employed if a very narrow

beam is required, the source being located either by measuring the variation in received signal strength as the earth rotates, or by electrically scanning the sky.

The interferometer[65] makes use of the interference pattern produced by two widely spaced aerial elements, and is analogous to the wide-aperture direction-finder. The principle is illustrated in Fig. 23: the broken line shows the radiation pattern of a

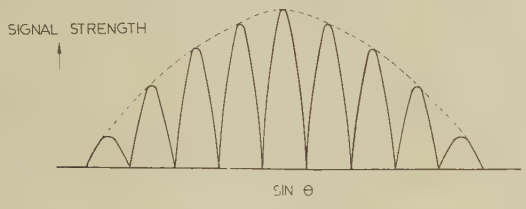


Fig. 23.—Principle of the interferometer.
---- Radiation pattern of single source.
—— Radiation pattern of two widely spaced sources.
θ = angle to normal

single element, and the solid line that for two widely spaced elements. By recording the received field strength as a function of time, the source can be located.

Two linear arrays in the form of a cross (the Mills cross[66]) may be used. By multiplying the outputs from the two aerials electronically, the effective radiation pattern for the cross can be made the product of the radiation patterns of the individual arms; in other words, a pencil-shaped beam is achieved. By swinging the phase of the currents in individual elements, the sky can be effectively scanned in azimuth or elevation. Side-lobe radiation may be minimized by grading the current distribution along each arm as described in Section 5.4. The arms may comprise any convenient radiating element, e.g. a dipole in association with either a parabolic cylinder or a small paraboloid. There is virtually no limit to the resolution achievable with such a system, but the sensitivity is limited by its relatively small collecting area.

An alternative procedure free from this disadvantage is 'aperture synthesis'.[67] This consists in measuring the relative amplitude and phase of the outputs of two small aerials occupying in turn all possible relative positions over the aperture, and then using a computer to calculate what the output would have been were the aperture completely filled; the effect of scanning can also be accomplished in the computation process. In the practical system a fixed east–west line of aerials extends over the length of the aperture, and a smaller aerial is moved over a north–south line extending over the width of the aperture.

(7.4) Supergain Aerials

A number of workers have studied arrays having an aperture of finite size but capable of giving an extremely sharp directional pattern. Such arrays can theoretically have unlimited gain and have therefore become known as supergain aerials. An aerial may be considered to exhibit supergain when the gain is higher than the largest obtainable if the elements are fed with equal currents, their contributions in the direction of the main lobe adding in phase. For the broadside array the normal gain is that achieved when the currents in the elements are equal and co-phased. For end-fire arrays the normal gain is that achieved when the currents are equal and in progressive phase. We can regard the degree of supergain as the excess of the actual gain over the normal gain. The possibility of supergain arrays was first envisaged by Oseen[68] in 1922. Schelkunoff[69] has analysed the case of broadside arrays and Goward[70] that of end-fire arrays. A comprehensive list of references on the subject has been given by Bloch *et al.*[71] Designing an aerial to have supergain is probably only worth while if the number of elements is comparatively small and the aerial is intended for metric or longer wavelengths, since only then is mechanical complication a limiting factor.

There is a limit to the achievable gain if the number of discrete sources is finite, but not if an infinite number is permissible, even though they may be restricted to an aperture of finite size. However, to achieve a high gain it is necessary for the current distribution to oscillate rapidly in sign across the aperture, with three consequent disadvantages. First, the radiation resistance is extremely low, and the losses are therefore so high that the theoretical gain is not realized; second, the bandwidth of the system is very low, and third, the currents in individual sources must be maintained to a high degree of accuracy. From published information on specific supergain arrays it appears that it is rather easier to achieve supergain with end-fire than with broadside arrays. Supergain is an interesting concept but it has not found any general practical application, although the Yagi aerial (Section 4.2) may be regarded as having a small degree of supergain.

(7.5) Data-Processing Techniques

Almost all the aerials described in this review work on the principle (when receiving) of taking a number of samples of the field over an aperture and transferring these to the aerial output terminals through linear invariant networks. In these cases the reciprocity principle applies and the performance characteristics relate to either transmission or reception. In a few instances, however, the aperture distribution is processed in some way before interpretation, e.g. in the multiplicative process used in the Mills cross (see Section 7.3). The individual samples of the incident field can be added or multiplied (with or without change of phase), they can be passed through linear, non-linear or time-varying networks, or be subject to combinations of any of these processes.[72] These techniques are at present considered applicable only to receiving aerials. Some of them are non-linear, and the reciprocity principle does not then apply.

(7.6) Wide-Band Aerials

It is sometimes desirable to provide an aerial whose impedance and radiation pattern remain the same over a band of many octaves. This is usually a requirement only for reception and when space is at a premium.

One way of satisfying it is to employ an aerial which has a form determined only by angles, e.g. the equiangular or logarithmic spiral[73]; an example is shown in Fig. 24. An alternative

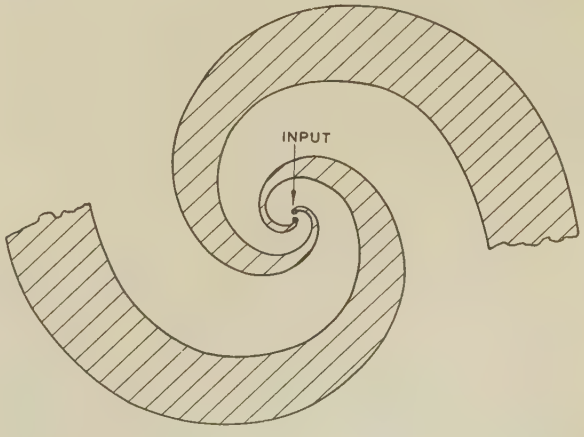


Fig. 24.—Equiangular-spiral aerial.

approach is to use an aerial whose electrical dimensions are the same at frequencies related by a geometric series; the performance then varies periodically with the logarithm of the frequency.[74] By making the periodicity (expressed as a function of frequency) sufficiently rapid, the performance can be made virtually independent of frequency. An example of a 'log-periodic' aerial is shown in Fig. 25.

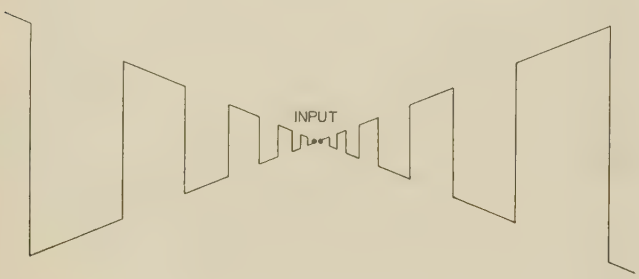


Fig. 25.—A log-periodic aerial.

To achieve the specified requirements over an infinite frequency band the aerial must extend to infinity, the effective radiating portion changing with the frequency. In practice the requirements can be fulfilled only over a limited band; the lower frequency limit is set by the maximum permissible size of the aerial, and the upper limit by the finite size of the input terminals.

(8) CONCLUSIONS

This review began by describing the aerial as a leaky waveguide. Since Hertz' first demonstration in 1888 enormous strides have been made in the understanding and control of the guiding process, but much still remains to be done in the field of theoretical analysis. Increasing attention is being given to the problem of radiation from apertures, and the related diffraction and scattering processes. This work is likely to have important repercussions on aerial design in the future, in providing exact solutions (or at least better approximations) to related radiation problems. Electronic computers are likely to be of great assistance in making use of solutions which it would be exceedingly laborious to deal with manually. Aerial design is likely to remain a fruitful field for research for a long time to come.

(9) ACKNOWLEDGMENTS

The author wishes to acknowledge his debt to fellow scientists, too numerous to mention individually, for their valuable advice concerning the presentation of the material contained in this review. He also wishes to thank the following organizations for permission to reproduce photographs: British Broadcasting Corporation [Figs. 1 and 11(*a*)], Radio Corporation of America [Figs. 11(*b*) and (*c*)], and Elliott Brothers (London) Ltd. [Fig. 18(*b*)]. The paper is published by permission of the Director of Engineering of the British Broadcasting Corporation.

(10) REFERENCES

* These papers are useful as sources of further references.

(1)*Ramsay, J. F.: 'Microwave Antenna and Waveguide Techniques before 1900', *Proceedings of the Institute of Radio Engineers*, 1958, **46**, p. 405.
(2)*Kiely, D. G.: 'Dielectric Aerials' (Methuen, 1953).
(3) Slee, J. A.: 'The Development of the Wireless Aerial', *Wireless World*, 1935, **37**, pp. 344, 368, 369, 468.
(4) Pierce, G. W.: 'Theoretical Investigation of the Radiation Characteristics of an Antenna', *Proceedings of the American Academy of Arts and Sciences*, 1916/17, **52**, p. 191.
(5) Ballantine, S.: 'High Quality Radio Broadcast Transmission and Reception', *Proceedings of the Institute of Radio Engineers*, 1934, **22**, p. 564.
(6) Franklin, C. S.: 'Short-Wave Directional Wireless Telegraphy', *Journal I.E.E.*, 1922, **60**, p. 930.
(7) 'Imperial Wireless "Beam" Communication', *Electrical Review*, 1926, **99**, pp. 709, 749.
(8) Harbich, H., and Hahnemann, W.: 'Wirksame Bekämpfung des Nahschwundes im Rundfunk durch Sendeantennengebilde bestimmter Form', *Elektrische Nachrichten-Technik*, 1932, **9**, p. 361.
(9) Chireix, H.: 'Antennes à rayonnement zénithal réduit', *L'Onde Électrique*, 1936, **15**, p. 440.
(10) Page, H.: 'Radiation Resistance of Ring Aerials', *Wireless Engineer*, 1948, **25**, p. 102, and 'Ring Aerial Systems', *ibid.*, p. 308.
(11) Magnusson, E., and Stranden, F.: 'Planning the New Motala Long-Wave Broadcasting Station', *European Broadcasting Union Review*, Part A, No. 61, 1960, p. 107.
(12) Brillouin, L.: 'Sur l'origine de la résistance de rayonnement', *Radioeléctricité*, 1922, **3**, p. 147.
(13)*Monteath, G. D.: 'The Effect of the Ground Constants, and of an Earth System, on the Performance of a Vertical Medium-Wave Aerial', *Proceedings I.E.E.*, Paper No. 279 R, January, 1958 (**105** C, p. 292).
(14) Stratton, J. A., and Chu, L. J.: 'Steady-State Solutions of Electromagnetic Field Problems', *Journal of Applied Physics*, 1941, **12**, p. 230.
(15)*Schelkunoff, S. A.: 'Advanced Antenna Theory' (Wiley and Sons, 1952), Chapter 2.
(16) Hallén, E.: 'Theoretical Investigations into the Transmitting and Receiving Qualities of Antennae', *Nova Acta* (Uppsala), 1938, **11**, Part 4.
(17)*King, R. W. P.: 'The Theory of Linear Antennas' (Harvard University Press, 1956), Chapter 2.
(18) Wells, N.: 'Aerial Characteristics', *Journal I.E.E.*, 1942, **89**, Part III, p. 76.
(19)*Page, H., and Monteath, G. D.: 'The Vertical Radiation Patterns of Medium-Wave Broadcasting Aerials', *Proceedings I.E.E.*, Paper No. 1714 R, September, 1954 (1955, **102** B, p. 279).
(20) Brown, G. H.: 'Directional Antennas', *Proceedings of the Institute of Radio Engineers*, 1937, **25**, p. 78.
(21) Walmsley, T.: 'Beam Arrays and Transmission Lines', *Journal I.E.E.*, 1931, **69**, p. 299.
(22) Yagi, H.: 'Beam Transmission of Ultra-Short Waves', *Proceedings of the Institute of Radio Engineers*, 1928, **16**, p. 715.
(23) Beverage, H. H., Rice, C. W., and Kellogg, E. W.: 'The Wave Antenna, a New Type of Highly Directive Antenna', *Transactions of the American I.E.E.*, 1923, **42**, p. 215.
(24) Foster, D.: 'Radiation from Rhombic Antennas', *Proceedings of the Institute of Radio Engineers*, 1937, **25**, p. 1327.
(25) Booth, C. F., and Maclarty, B. N.: 'The New High-Frequency Transmitting Station at Rugby', *Proceedings I.E.E.*, Paper No. 1903 R, October, 1955 (1956, **103** B, p. 263).
(26)*Norman, F. J., and Ward, J. F.: 'Rhombic Aerials', *Electronic and Radio Engineer*, 1957, **34**, p. 398.
(27) Friis, H. T., and Feldman, C. B.: 'A Multiple Unit Steerable Antenna for Short-Wave Reception', *Proceedings of the Institute of Radio Engineers*, 1937, **25**, p. 841.
(28) Morris, D. W., and Mitchell, G.: 'A Multiple-Direction Universally Steerable Aerial System for H.F. Operation', *Proceedings I.E.E.*, Paper No. 3078 E, November, 1959 (**106** B, p. 555).
(29) Roberts, W. van B.: 'Input Impedance of a Folded Dipole', *RCA Review*, 1947, **8**, p. 289.
(30) Kraus, J. D.: 'Antennas' (McGraw-Hill, 1950), Chapter 7.
(31) Booker, H. G.: 'Slot Aerials and Their Relation to Complementary Wire Aerials (Babinet's Principle)', *Journal I.E.E.*, 1946, **93**, Part IIIA, p. 620.
(32)*Wait, J. R.: 'Electromagnetic Radiation from Cylindrical Structures' (Pergamon Press, 1959).
(33) Carter, P. S.: 'Antenna Arrays Around Cylinders', *Proceedings of the Institute of Radio Engineers*, 1943, **31**, p. 671.
(34) Knight, P.: 'Methods of Calculating the Horizontal Radiation Patterns of Dipole Arrays Around a Support Mast', *Proceedings I.E.E.*, Paper No. 2740 R, November, 1958 (**105** B, p. 548).
(35) Dolph, C. L.: 'A Current Distribution for Broadside Arrays which Optimizes the Relationship Between Beam Width and Side-Lobe Level', *Proceedings of the Institute of Radio Engineers*, 1946, **34**, p. 335.
(36) Woodward, P. M.: 'A Method of Calculating the Field Over a Plane Aperture Required to Produce a given Polar Diagram', *Journal I.E.E.*, 1946, **93**, Part IIIA, p. 1554.
(37) Ksienki, A.: 'Derivative Control in Shaping Antenna Patterns', *Institute of Radio Engineers International Convention Record*, 1960, **8**, Part 1, p. 3.
(38)*Harvey, A. F.: 'Optical Techniques at Microwave Frequencies', *Proceedings I.E.E.*, Paper No. 2779 E, March, 1959 (**106** B, p. 141).
(39) Booker, H. G., and Clemmow, P. C.: 'The Concept of an Angular Spectrum of Plane Waves, and its Relation to that of Polar Diagram and Aperture Distribution', *ibid.*, Paper No. 922, January, 1950 (**97**, Part III, p. 11).
(40) Cottony, H. V., and Wilson, A. C.: 'Gains of Finite-Size Corner Reflector Antennas', *Transactions of the Institute of Radio Engineers*, **AP–6**, 1958, p. 366.
(41) Wilson, A. C., and Cottony, H. V.: 'Radiation Patterns of Finite Size Corner-Reflector Antennas', *ibid.*, **AP–8**, 1960, p. 144.
(42)*Medhurst, R. G.: 'Passive Microwave Mirrors', *Electronic and Radio Engineer*, 1959, **36**, p. 443.

(43) BEDROSIAN, E.: 'The Curved Passive Reflector', *Transactions of the Institute of Radio Engineers*, **AP–3**, 1955, p. 168.
(44) KRAUS, J. D., NASH, R. T., and KO, H. C.: 'Some Characteristics of the Ohio State University 360-ft Radio Telescope', *ibid.*, **AP–9**, 1961, p. 4.
(45) KOCK, W. E.: 'Metallic Delay Lenses', *Bell System Technical Journal*, 1948, **27**, p. 58.
(46) BROWN, J., and JACKSON, W.: 'The Properties of Artificial Dielectrics at Centimetre Wavelengths', *Proceedings I.E.E.*, Paper No. 1699 R, January, 1955 (**102** B, p. 11).
(47) KOCK, W. E.: 'Path-Length Microwave Lenses', *Proceedings of the Institute of Radio Engineers*, 1949, **37**, p. 852.
(48) SIMON, J. C.: 'Un nouveau type de lentilles en hyperfréquences', *L'Onde Électrique*, 1952, **32**, p. 181.
(49) RUST, N. M.: 'The Phase Correction of Horn Radiators', *Journal I.E.E.*, 1946, **93**, Part IIIA, p. 50.
(50) PIPPARD, A. B.: 'The Hoghorn—An Electromagnetic Horn Radiator of Medium-Sized Aperture', *ibid.*, 1946, **93**, Part IIIA, p. 1536.
(51) CORBIN, A. T., and MAY, A. S.: 'Broadband Horn Reflector Antenna', *Bell Laboratories Record*, 1955, **33**, p. 401.
(52) FRY, D. W.: 'Slotted Linear Arrays', *Journal I.E.E.*, 1946, **93**, Part IIIA, p. 43.
(53) BROWN, S. G.: British Patent No. 14449, 1899.
(54) ROUND, H. J.: 'Direction and Position Finding', *Radio Review*, 1920, **1**, pp. 235 and 289.
(55)*SMITH-ROSE, R. L., and BARFIELD, R. H.: 'The Cause and Elimination of Night Errors in Radio Direction-Finding', *Journal I.E.E.*, 1926, **64**, p. 831.
(56) EARP, C. W., and GODFREY, R. M.: 'Radio Direction-Finding by the Cyclical Differential Measurement of Phase', *ibid.*, 1947, **94**, Part IIIA, p. 705.
(57)*HOPKINS, H. G., and PRESSEY, B. G.: 'Current Direction-Finding Practice', *Proceedings I.E.E.*, Paper No. 2579 R, March, 1958 (**105** B, Suppl. No. 9, p. 307).
(58) RATCLIFFE, J. A.: 'Aerials for Radar Equipment', *Journal I.E.E.*, 1946, **93**, Part IIIA, p. 22.
(59) FRIIS, H. T., and LEWIS, W. D.: 'Radar Antennas', *Bell System Technical Journal*, 1947, **26**, p. 219.
(60) BROWN, J.: 'Microwave Lenses' (Methuen, 1953), Chapter 6.
(61) LUNEBERG, R. K.: 'Mathematical Theory of Optics' (Brown University Press, 1944), p. 212.
(62) PEELER, G. D. M., and COLEMAN, H. P.: 'Microwave Stepped-Index Luneberg Lenses', *Transactions of the Institute of Radio Engineers*, **AP–6**, 1958, p. 202.
(63) JONES, S. S. D.: 'A Wide-Angle Microwave Radiator', *Proceedings I.E.E.*, Paper No. 988, July, 1950 (**97**, Part III, p. 255).
(64) PEELER, G. D. M., and ARCHER, D. H.: 'A Two-Dimensional Microwave Luneberg Lens', *Transactions of the Institute of Radio Engineers*, **AP–1**, 1953, p. 12.
(65) RYLE, M.: 'A New Radio Interferometer and its Application to the Observation of Weak Radio Stars', *Proceedings of the Royal Society*, 1952, **211** A, p. 351.
(66) MILLS, B. Y., and LITTLE, A. G.: 'A High-Resolution Aerial System of a New Type', *Australian Journal of Physics*, 1953, **6**, p. 272.
(67) RYLE, M.: 'The Mullard Radio Astronomy Observatory', *Journal I.E.E.*, 1960, **6**, p. 14.
(68) OSEEN, C. W.: 'Die Einsteinsche Nadelstichstrahlung und die Maxwellschen Gleichungen', *Annalen der Physik*, 1922, **69**, p. 202.
(69) SCHELKUNOFF, S. A.: 'A Mathematical Theory of Linear Arrays', *Bell System Technical Journal*, 1943, **22**, p. 80.
(70) GOWARD, F. K.: 'An Improvement in End-Fire Arrays', *Journal I.E.E.*, 1947, **94**, Part III, p. 415.
(71)*BLOCH, A., MEDHURST, R. G., and POOL, S. D.: 'Superdirectivity', *Proceedings of the Institute of Radio Engineers*, 1960, **48**, p. 1164.
(72) Unpublished papers presented at 13th General Assembly of Union Radio Scientifique Internationale, London, 1960.
(73) BAWER, R., and WOLFE, J. J.: 'The Spiral Antenna', *Institute of Radio Engineers International Convention Record*, Part 1, 1960, p. 84.
(74) DUHAMEL, R. H., and ORE, F. R.: 'Logarithmically Periodic Antenna Designs', *Institute of Radio Engineers National Convention Record*, Part 1, 1958, p. 139.

621.317.36.029.5 : 621.372.824

The Institution of Electrical Engineers
Paper No. 3691 M
Sept. 1961

PRECISION INSTRUMENTS FOR COAXIAL LINE MEASUREMENTS UP TO 4 Gc/s

By D. WOODS, Associate Member.

(*Lecture delivered before a joint meeting of the* MEASUREMENT AND CONTROL SECTION *and the* ELECTRONICS AND COMMUNICATIONS SECTION *11th April*, 1961.)

An unsatisfactory situation exists in this country arising from the lack of national radio-frequency standards, and Government organizations have been forced to take extraordinary measures in order to ensure that equipment accepted for Service use will meet the operational requirements. The provision of such standards is necessary in order to ensure, first, interoperability of equipment between the Services and N.A.T.O. countries, and secondly, reliable and indisputable performance of measuring instruments for the home and export market. The provision of national standards alone is not sufficient; they must be supplemented by a service for the calibration of standards held by Government and commercial organizations together with supporting advice on measuring techniques generally.

In 1948 there were no plans for the provision of adequate national facilities, so a long-term programme was initiated at the Government laboratories, Harefield, to effect an improvement, by a factor of at least ten, in the standardization and measurement of admittance, voltage, power and attenuation. The outcome of that programme is the main subject of the lecture.

The basic principle adopted for the programme was the realization of a range of wide-band instruments of such good voltage-standing-wave ratio (v.s.w.r.) that the effect of mismatch could be substantially ignored. Clearly, this placed the emphasis on admittance standardization because an improvement of nearly two orders of magnitude over current techniques was needed. This raised the question of suitable coaxial connectors of sufficiently low v.s.w.r. to ensure compatibility with the required accuracy on the admittance parameter. References 1 and 2 describe a precision dual admittance bridge, having an absolute inaccuracy of 0·2% up to 200 Mc/s, and Reference 3 describes a coaxial connector system, having a residual v.s.w.r. of 0·999 up to 4 Gc/s. The coaxial connector system has been adopted for all the instruments subsequently developed. For the range 400 Mc/s–4 Gc/s a precision slotted line, having an inaccuracy of 0·5–1·0%, was used. It is proposed to develop a new dual admittance bridge to cover the gap 200–500 Mc/s with an inaccuracy of 0·2%.

The next step in the programme was the development of a precision wide-band coaxial resistor mount which forms an essential part of terminating resistors, dissipative attenuators, voltmeters, wattmeters, insertion-loss meters, etc. The mount employs a cylindrical film resistor of uniform surface resistivity surrounded by a metallic outer conductor having a tractorial profile.[4,5] A typical 50 Ω terminating resistor has a v.s.w.r. of 0·99 from d.c. to 4 Gc/s. The principle of operation is described in terms of electromagnetic-wave propagation over a resistive film surface in which the tilt angle θ of the electric field from the normal to the film surface is given by $\sin\theta = \rho/377$, where ρ is the surface resistivity. Attention is drawn to the importance of using a properly designed coaxial conical line and reflectionless support to connect the tractorial mount to the coaxial cylindrical line of the input connector.

Turning to the subject of voltage and power measurement, I shall now discuss the various ways of specifying or calibrating the output level of a nominally matched source. The advantages of the concepts of equipment source e.m.f. or equivalent available power are given in Reference 6. These concepts render the source output-level characteristic substantially independent of its impedance and place the requirement for accurate matching on the measuring instrument. Furthermore, the two concepts are compatible in that a source calibrated by either method will lead to the same estimate of power dissipated in any load impedance. Because the measuring instrument has to be accurately matched this means that it can be designed to perform the dual function of voltage and power measurements, so that these two quantities are related by $P = V^2G$. Three instruments are designed for this purpose:

(*a*) A coaxial crystal millivoltmeter/milliwattmeter,[7,8] 50 or 75 Ω, covering all radio frequencies up to 1 Gc/s with a v.s.w.r. of 0·995 and an inaccuracy of 0·5% on voltage and 1·5% on power. This instrument has a direct-reading power scale linear to 0·1% and is fully compensated to within 0·1% for an ambient-temperature range 10–30° C. The crystal head is illustrated in Fig. 1. The instrument is calibrated at 1 Mc/s by means of a

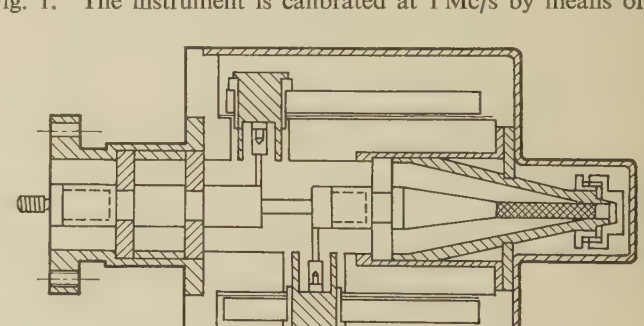


Fig. 1.—Section of crystal milliwattmeter head.

d.c.–1 Mc/s transfer standard having an inaccuracy of 0·25% on voltage at 1 Mc/s.

(*b*) A coaxial film bolometer,[9] 50 Ω, covering 200 Mc/s–4 Gc/s with a v.s.w.r. of 0·98 and an inaccuracy of 1% between 20 and 200 mW.

(*c*) A thermistor milliwattmeter, 50 Ω, covering 100–1 250 Mc/s with a v.s.w.r. of 0·97 and an inaccuracy of 1% between 20 μW and 2 mW.

The three instruments, each calibrated in terms of d.c. standards, have been compared in the u.h.f. band and the agreement obtained was between 1–2% in power. A fourth instrument has been developed for the measurement of peak pulse power. This covers the range 100–1 250 Mc/s with a v.s.w.r. of 0·98 and an inaccuracy of 2% for duty cycles down to 0·001 in the power range 0·01–0·1 W. The importance of the admittance parameter on the accuracy of terminated voltmeters and wattmeters is illustrated in Fig. 2. The curves are drawn to give an error of ±1% for a source v.s.w.r., S_s, of 0·5. This condition arises when the voltmeter v.s.w.r., S_v, is 0·985 and the wattmeter v.s.w.r., S_w, is 0·97. The error is critically dependent on S_v even for a perfectly matched source. This condition also applies

Mr. Woods is at the Ministry of Aviation.

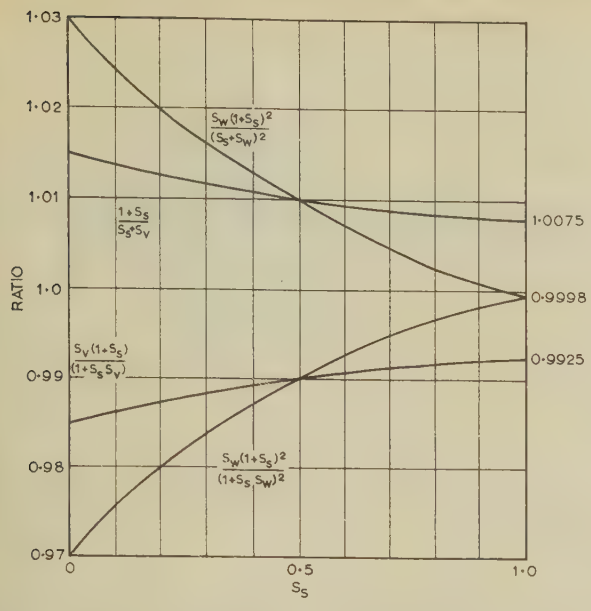


Fig. 2.—Ratio of the indications of imperfect/perfect instruments.

to wattmeters which operate on the principle of voltage squared times the nominal conductance, but not to those which operate on the thermal principle and to which the S_w curves of Fig. 2 apply.

The measurement of intermediate and high levels of voltage and power was carried out with the aid of dissipative attenuators or directional couplers used in conjunction with the instruments previously described. The dissipative attenuators employ tractorially mounted resistors in an L-network and cover up to 50dB in steps of 5dB. They have a frequency range from d.c. to several gigacycles per second, a v.s.w.r. of 0·98 and an inaccuracy of 0·1dB. A 20dB 20W dissipative attenuator is illustrated in Fig. 3. The directional couplers are based on the principle described by Monteath.[10] They comprise two precision 50Ω coaxial lines mutually coupled over a short distance by means of a knife-edge axial slot in the outer conductors.

Fig. 3.—Dissipative attenuator: 50Ω, 20dB, 20W.

The attenuation is a function of the electrical length of the slot, being a minimum when this is an odd number and infinite when it is an even number of quarter wavelengths. The width of the slot governs the minimum attenuation. Owing to the attenuation/frequency characteristic it is important that the source is free from harmonics, otherwise errors can be caused by harmonics being less attenuated than the fundamental. A unique property of the directional coupler is that two wattmeters can be compared, one in the primary and one in the secondary line, under conditions which are independent of the source impedance provided that the directivity is good. If the directivity is poor, this gives rise to an uncertainty factor in the comparison of the two wattmeters. A curve is shown in Fig. 4 which gives the limits of this uncertainty in terms of the

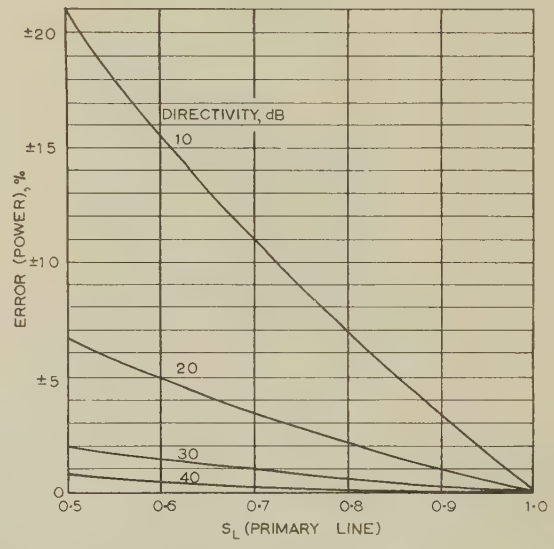


Fig. 4.—Errors caused by directivity.

v.s.w.r., S_L, of the primary-line wattmeter and the directivity. It is apparent that for precise work a directivity of at least 30dB is necessary. Three pairs of couplers have been designed for nominal attenuations of 10, 20 and 30dB. Each pair has different lengths of slot so that the attenuation/frequency characteristics overlap in such a manner that a useful range of attenuation is provided between 150Mc/s and 4Gc/s for each value of nominal attenuation. The couplers were designed to within 0·1dB but they could be calibrated to a higher order of accuracy because their stability is 0·02dB.

Attenuation measurement and its two forms, incremental attenuation and insertion loss, will next be discussed. Measurement of the former is substantially independent of the source and load impedance characteristics provided that a sufficient degree of isolation is present. This is not so in the latter case where the loss arising through the insertion of the attenuator into a coaxial system is a function of five complex quantities,[11] namely the reflection coefficients of the source and load and the input and output to the attenuator together with its voltage transfer coefficient. If 'insertion loss' is defined as the attenuation arising when the source and load are resistive and equal to a specified nominal characteristic impedance (such as that of the coaxial connectors attached to the attenuator), the insertion loss is equal to the reciprocal of the modulus of the voltage transfer coefficient alone. This is a very convenient definition of insertion loss because it depends on only one parameter of the attenuator. Furthermore, it is independent of the direction of transmission through the attenuator provided that it is a passive

reciprocal device. The measurement of insertion loss, defined in this manner, requires a well-matched source and load. Signal generators and attenuation-measuring receivers do not normally present very good matches. To overcome this a pair of source and load units, employing tractorially mounted resistors, have been designed. These are interposed between the attenuator and the poorly-matched source of power and attenuation-measuring receiver. The attenuation-measuring receiver (Fig. 5)

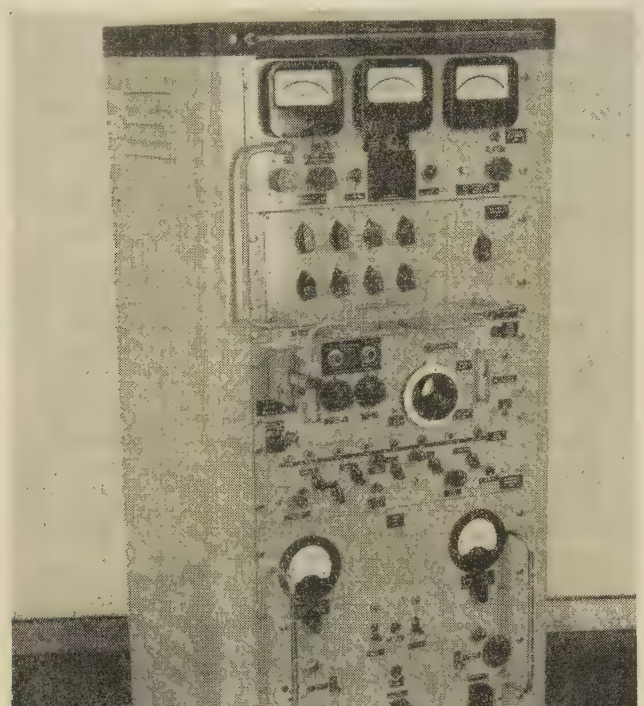


Fig. 5.—Attenuation-measuring receiver.

covers the ranges 50 kc/s–200 Mc/s and 0–126 dB with an inaccuracy of 0·05 dB above the 10 μV level and 0·1 dB between 1 and 10 μV. The discrimination is 0·01 dB. The equipment operates at an intermediate frequency of 1 Mc/s and this channel contains the standard attenuator. This is a resistive ladder network continuously adjustable from 1 to 127 dB in steps of 0·1 dB with an inaccuracy of 0·01 dB.

Noise-factor measurements are becoming increasingly important in view of recent developments in the field of parametric amplifiers and masers as applied to space communication and radio astronomy. The causes of the relatively poor performance of currently-available coaxial noise sources are due to poor source matching and errors caused by transit-time effect and incipient resonance of the noise diode. A section of the head of a new noise source of improved performance is shown in Fig. 6. It employs a tractorially mounted 50 Ω resistive load and a pair of specially developed diodes (type E 2790) connected across the coaxial line. The diode capacitances are neutralized by means of an undercut in the inner conductor. It has a frequency range of 30–1 250 Mc/s with a v.s.w.r. of 0·97. The inaccuracy is 0·1 dB up to 200 Mc/s, increasing to 0·25 dB at 1 250 Mc/s for noise factors up to 20. Improvement in the accuracy at the

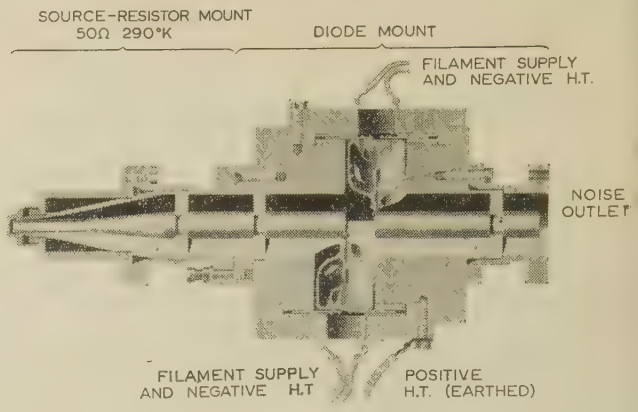


Fig. 6.—Noise generator.

higher frequencies is possible by the application of corrections. Plans are in hand to assess the corrections by means of a standard hot load.

Acknowledgments are due to the staff at the E.I.D. Laboratories, Harefield, for their contributions in the execution of the development programme; in particular to Mr. I. A. Harris, who worked closely with the author whilst he was at the Laboratories and who has been responsible for the continuation and extension of the programme since the author left. Special mention should be made of Mr. Harris's invention of the tractorial resistor mount. Acknowledgments are also due to Mr. R. W. A. Siddle for his work on the attenuation measuring receiver, the 1 Mc/s calibrator and supporting development on associated electronic equipment; to Mr. L. J. Graver for calculations connected with the noise source and directional couplers; and to Mr. R. E. Spinney for extensive and precise measurements in support of the programme generally.

REFERENCES

(1) Woods, D.: 'A Precision Dual Bridge for the Standardization of Admittance at Very High Frequencies', *Proceedings, I.E.E.*, Monograph No. 244R, June, 1957 (**104** C, p. 506).
(2) Woods, D.: U.K. Patent No. 681927, October, 1952.
(3) Woods, D.: 'A Coaxial Connector System for Precision R.F. Measuring Instruments and Standards', *Proceedings I.E.E.*, Paper No. 3499E, March, 1961 (**108**B, p. 205).
(4) Harris, I. A.: 'The Theory and Design of Coaxial Resistor Mounts for the Frequency Band 0–4 000 Mc/s', *ibid.*, Monograph No. 132R, May, 1955 (**103**C, p. 1).
(5) Harris, I. A.: U.K. Patent No. 670339, April, 1952.
(6) Woods, D.: 'The Concept of Equivalent Source E.M.F. and Equivalent Available Power in Signal-Generator Calibration', *Proceedings I.E.E.*, Paper No. 3356M, January, 1961 (**108**B, p. 37).
(7) Woods, D.: 'A Coaxial Millivoltmeter/Milliwattmeter for Frequencies up to 1 Gc/s', *ibid.* (to be published).
(8) Woods, D.: U.K. Patent No. 842720, July, 1960.
(9) Harris, I. A.: 'A Coaxial Film Bolometer for the Measurement of Power in the U.H.F. Band', *Proceedings I.E.E.*, Paper No. 3146E, January, 1960 (**107**B, p. 67).
(10) Monteath, G. D.: 'Coupled Transmission Lines as Symmetrical Directional Couplers', *ibid.*, Paper No. 1833R, May, 1955 (**102**B, p. 383).
(11) Beatty, R. W.: 'Mismatch Errors in the Measurement of Ultra High Frequency and Microwave Variable attenuators', *Journal of Research of the National Bureau of Standards*, January, 1954, **52**, p. 7.

DISCUSSION BEFORE A JOINT MEETING OF THE MEASUREMENT AND CONTROL SECTION AND THE ELECTRONICS AND COMMUNICATIONS SECTION, 11TH APRIL, 1961

Dr. J. Brown: The accuracy with which measurements could be made in the frequency range 300–3 000 Mc/s was an order of magnitude less than is now possible with the improved techniques described by the author. The difficulties in this range arise because the techniques involved are a mixture of conventional circuit design and waveguide plumbing. The illustrations indicate that the author has biased his designs towards using waveguide techniques, and there are close similarities between several of his components and the corresponding waveguide ones. The couplings between the coaxial sections are very similar to waveguide couplings, and the directional-coupler design is comparable to a standard microwave type. This similarity between the techniques used by the author and those used for microwave measurements suggests that a brief comparison would be valuable.

Voltage measurements cannot be made for microwave signals and so no comparison can be made. As far as power measurements are concerned, the methods are virtually the same. Instruments based on thermal effects can be used from direct current up to 10 Gc/s and higher frequencies with an accuracy of 1%. A considerable difference arises in the method used for impedance measurements. The author has used a bridge method throughout the whole frequency range, whereas microwave measurements rely on a standing-wave indicator. He mentions the coaxial-line standing-wave indicator but gives the impression that the bridge was more accurate. Is this the case, even towards the high-frequency end of the range considered?

Impedance measurements can also be made by directional couplers. A directivity of 40–50 dB is needed to give an accuracy comparable with that of the author's bridge, and this is quite possible at microwave frequencies. For frequencies of 300–400 Mc/s upwards it should be possible to produce directional couplers of this performance, possibly at the expense of making them longer. A different approach may then be used in which the directional coupler forms the bridge network. This would have many advantages. It would be more convenient to handle than a coaxial standing-wave indicator. Standing-wave measurements are tedious at best, and they are subject to a fairly large number of errors, as is pointed out in Reference 6. Is any work in progress on such directional couplers?

One interesting difference which arises is in nomenclature. The term 'precision measurement' in the microwave range tends to be used for a class of measurement techniques. These techniques are based on the idea that, if one can take a series of observations, draw a curve of theoretically known shape and calculate the required information from the parameters of this curve, the result will presumably be more accurate than relying on one observation. The author is using precision in the sense of B.S.I. instruments, which implies an accuracy of about 0·5%. The microwave type of precision measurement raises the possibility of further development in the author's field, and it would be particularly useful in connection with attenuation measurements. The measurement of waveguide components with insertion losses of up to 10 dB can be made extremely accurately by backing the device with a sliding short-circuit and making measurements of the impedance of the combination for a succession of short-circuit positions. This method is capable of very high accuracy, not only for the insertion loss itself but also for the reflection coefficients of the component. The measured points theoretically should lie on a circle. The centre of the circle, its radius and another easily defined point can be measured and used to give a complete specification of the component behaviour at the measurement frequency.

One of the most interesting contributions is the treatment of the mismatch errors which are associated with generators. The author's suggestion for changing the method of specifying the behaviour of a single generator is very useful. The full account, which is given in Reference 6, shows clearly that the concept of an equivalent source e.m.f. leads to a considerable reduction in the possible errors which can arise. Is it possible that a similar method can be applied to noise generators?

The tractrix which has been used for the coaxial resistor mounts intrigues me. The simple explanation given is convincing as showing that it is likely to give a successful result, but there is a danger in using what is essentially an optical argument and applying it to a system in which the dimensions are extremely small compared with wavelength. In this case it is true that the results have been confirmed by a more exact theoretical approach. Has it been confirmed experimentally that the tractrix does give the optimum performance, or whether one can produce equivalent results by a similar type of curve?

Mr. E. M. Lee: Is it correct that these connectors are a hundred times more accurate than existing types? The author has shown us the effect of the length of the slot in the directional couplers. Is the width of the slot critical and, if so, how critical?

Mr. L. Lewin: Why has the size of the cable been designed in inches? The coaxial-line figure was 1·3 in. Why not choose 1¼ in? For scientific instruments centimetre measurements are desirable, especially when comparisons are made with international standards. It is difficult to construct coaxial lines without dielectric support. How were the inner conductors supported and what were the frequency variations arising therefrom?

Has any difficulty been experienced from contacts? My experience of the problem has been in waveguides, but I imagine that the same trouble arises in coaxial lines, particularly at higher frequencies. I gathered from the way in which the pieces were connected that one had a butt joint. Quite large currents can flow in the gap, giving resistivity losses and reactive discontinuities. This is a drawback for the accuracies the author is considering. One can combat it with shims, and there are other ways.

I infer that the attenuation in the directional couplers is frequency sensitive. The author mentions the effect of width as altering attenuation levels; presumably it is also frequency sensitive. On the question of directivity, it is a little unexpected that a slot will couple *all* power in the backward direction and none in the forward direction. This must depend on a balance of electric and magnetic coupling. This would be affected both by the width and by the length of the slot.

Mr. A. C. Lynch: Does the author see any hope of making variable resistors for these frequencies, and thus widening the range of bridge circuits that can be realized?

Mr. D. Woods (*in reply*): Dr. Brown refers to the similarity between the coaxial components and the corresponding waveguide ones. This similarity is closest in the design of the butt-flange connectors and the tight mechanical tolerances needed. The realization of a coaxial component of comparable performance to a waveguide component is, generally speaking, more difficult owing to the presence of the inner conductor. The bandwidth of coaxial components can be made much wider than waveguide ones, i.e. from d.c. to several gigacycles per second, and furthermore, the impedance can be made independent of frequency over this range. With waveguides the lower frequency is set by the cut-off frequency and the higher frequency by the

onset of higher modes. The useful frequency range is generally limited to a smaller value by other factors such as stub matches and sliding short-circuits. Attention is drawn to the fact that these artifices are not used in the coaxial instruments described because their absence is an essential feature in a wide-band device.

For admittance measurement the bridge method is used from 3 to 250 Mc/s with an inaccuracy of 0·1–0·2%. A slotted line is used from 400 to 4 000 Mc/s because there is no alternative. This has an inaccuracy of 0·5–1% and is the best that can be realized with current manufacturing methods. Nevertheless, higher accuracy is needed in this range. The highest frequency for the bridge method is considered to be 500 Mc/s, and plans are in hand to develop an instrument of comparable performance to the existing bridge. The possibility of using directional couplers for more accurate admittance measurement, in place of the slotted line, has not been overlooked, but so far no work has been put in hand. If this method is to yield an accuracy comparable with that of the bridge technique then it is considered that a directivity very much better than 40–50 dB will be needed. It is inevitable that when a particular technique is adapted for measurements of greater accuracy many sources of error, hitherto negligibly small or unsuspected, becomes significant.

The use of the term 'precision measurement' in the microwave field to denote a class of measurement technique is noted with interest. This technique is also used in the calibration of some of the coaxial instruments described, in particular the derivation of the residual parameters of the dual admittance bridge (Ref. 2 of my Ref. 1). The equivalent waveguide technique could also be used for the measurement of small insertion losses. Because the attenuation measuring receiver meets the current requirement in the range 0·01–126 dB the alternative technique has not been adopted. There would be mechanical and electrical difficulties in providing a substantially loss-free coaxial sliding short-circuit for the lower frequencies.

It is doubtful whether the concept of equivalent source e.m.f. could be applied to noise generators because the error in noise-factor measurement depends on a number of noise parameters of the receiver in addition to the matching of the generator and the receiver to a nominal characteristic impedance. Furthermore, optimum noise factor does not coincide with optimum matching of the receiver to the source.

The tractrix is the ideal curve for the outer jacket of a coaxially mounted cylindrical film resistor of uniform and low surface resistivity. It also mates with the lossless conical input line with no discontinuity. The first experimental 25 Ω resistor mount was measured on the dual admittance bridge and it exhibited an unaccountable shunt capacitance (less than 0·1 pF) at 200 Mc/s. A fuller theory employing Maxwell's equations was subsequently developed[4] which took into account the penetration of the field into the resistor body. The theory showed that the field penetration caused a small distributed negative inductance throughout the length of the resistor [eqn. (87) of Reference 4]. In practice this is compensated by increasing the radius of the tractrix by a small fixed amount. This causes a negligibly small discontinuity between the tractrix and the input-cone outer conductor. The nearest approximation to the tractrix is the exponential curve, but this does not produce such a good performance as a curve of the former type.

Mr. Lee is correct in assuming that the connectors are between one and two orders of magnitude more accurate than existing types. For this reason coaxial cables cannot be used because of the relatively large discontinuities they produce. The attenuation of the directional couplers is critically dependent on the slot width. For example, the slot width of the 20 dB coupler is 0·43 in and a variation of 0·0016 in would produce a change in attenuation of about 0·1 dB. These figures relate to 50 Ω 0·750 in-outer-diameter primary and secondary lines.

In answer to Mr. Lewin, the slotted line, having an outer diameter of about 1·3 in, was constructed to a design existing before the 0·75 in connector was adopted. The latter figure is governed by the maximum frequency chosen, namely 4 Gc/s. The nearest metric size is 2 cm. If this had been adopted, the required diameter ratio of 1·5 for the bridge terminal and coaxial standards would make the inner conductor diameter 1·333 cm. By using inches the diameters are 0·750 in and 0·500 in and this had the advantage that precision-drawn silver tubes for the standards were available at the time the bridge was under development. The dielectric supports are not considered as part of the connector system, for the reasons given in Ref. 3. For the associated instruments a p.t.f.e. disc support is used with the inner and outer conductors undercut to compensate the discontinuity capacitances. A 50 Ω support, 0·15 in thick, has a residual v.s.w.r. of not less than 0·997 up to 4 Gc/s. The source of the design information is given in Ref. 3.

Great attention was paid to contacts when the bridge and connectors were under development. The variable capacitors in the bridge can be set to 1 part in 10^5 at several hundred megacycles per second without any difficulty. The flanges of the connectors are accurately machined to within about 0·0001 in and are gold plated. The contact resistance is of the order of 50 μΩ for the flange and 200 μΩ for the inner conductor. This corresponds to the resistance of a 50 Ω 0·750 in-outer-diameter silver line, 0·05 in long, at 200 Mc/s. This is negligibly small in relation to admittance measurements of 0·1% inaccuracy.

The attenuation/frequency curve of the directional couplers is a series of half sine waves with minimum attenuation occurring at the crests when the slot length is an odd number of quarter-wavelengths. The attenuation approaches infinity at the even number of quarter-wavelength points. The slot width governs the general level of the attenuation and is frequency independent when it is small compared with the wavelength. The field in the primary line couples into the secondary line through the slot. If the magnetic and electric fields in the former are, say, clockwise and from inner to outer, respectively, then the directions in the secondary line will be clockwise and from outer to inner, i.e. propagation in the secondary line will be in the opposite direction. Propagation in the same direction can be caused only by reflections from the slot ends and mismatches in the secondary line caused by, for example, dielectric supports and terminations. In a well-designed coupler it is the reflections from the slot ends that are most difficult to suppress for good directivity. This form of coupler is more appropriately called a contra-directional coupler. The theory is given in Reference 10.

Mr. Lynch's question leaves one in no doubt that he is thoroughly familiar with the problem of realizing a suitable circuit for a precision impedance bridge for radio frequencies, as pointed out in the conclusions of Reference 3. Indeed, many other useful bridge circuits would become available if a precision variable r.f. resistor was a practicable proposition. In order to achieve an impedance bridge of comparable accuracy to the dual admittance bridge described it would be necessary to design a variable resistor with a discrimination of about 2 parts in 10 and with an absolute inaccuracy of 0·1% over a range of at least 10 : 1 at several hundred megacycles per second. I see no hope of meeting this requirement in the light of the present state of the art. This means that r.f. bridges of this accuracy will have to rely for the time being on two variable reactances for balancing the in-phase and out-of-phase components.

621.396.93 : 621.396.67

The Institution of Electrical Engineers
Paper No. 3678 E
Sept. 1961

THE APPLICATION OF THE INTERFEROMETER TO H.F. DIRECTION-FINDING

By C. W. McLEISH, M.Sc., and N. BURTNYK, B.Sc.

(*The paper was first received 24th February, and in revised form 18th May*, 1961.)

SUMMARY

The measured performance of an interferometer system for direction-finding in the h.f. band is described. A considerable reduction in bearing error compared with a narrow-aperture (Adcock) system is obtained in the presence of ground reradiators and fluctuating wave-interference fields. A standard deviation of about 1·2° has been obtained with a 400 ft interferometer on sky-wave transmissions in the 5–25 Mc/s band.

(1) INTRODUCTION

Radio direction-finding in the ordinary sense implies finding the direction of discrete transmission sources. The measurement is based on finding either the mean direction of arrival of wave energy or the direction of arrival of the predominant mode. Two errors appear, one due to the uncertainty of determining the direction of arrival on the above basis, and the other to path deviations which make the direction of arrival differ from the true direction of the transmitter.

In the h.f. band (3–30 Mc/s) deviations of long-distance paths are usually correlated over large distances (tens of kilometres) so that it is impractical to consider means of reducing this error within a single direction-finder. On the other hand, the interference effects of angular dispersion of the received energy are correlated over distances proportional to wavelength and the cosecant of half the angular spread of interfering waves. Therefore it is usually possible to reduce wave-interference error by sampling the field at two or more points separated in space by several wavelengths. The theoretical performance of wide-aperture systems in wave-interference fields is discussed by Bain.[1]

A further consideration is the effect of time variations on the interference field and on the path deviations. Although this is negligible for fixed ground-wave paths, time variation may be important for ionospherically reflected paths. In the latter case, substantial reduction in wave-interference error can usually be obtained in times of the order of a minute, whereas path-deviation errors are sometimes correlated over periods of hours. Therefore in designing a direction-finding system, one should consider the relative advantages of antenna aperture and time of observation in the reduction of wave-interference errors.

The purpose of the paper is to show experimentally the performance of a 2-element interferometer for ground- and sky-wave reception and to compare it with the performance of a narrow-aperture Adcock system. In the sky-wave test, lateral deviation and wave-interference errors are separated statistically and the relative advantage of the interferometer over the Adcock is demonstrated. Section 6 contains a brief discussion on the problem of ambiguity resolution of the multi-lobed pattern of the interferometer.

(2) EXPERIMENT

(2.1) Equipment

An array configuration shown in Fig. 1 is connected in such a way as to measure sequentially the phase differences between

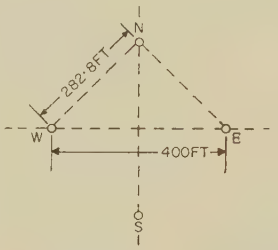


Fig. 1.—Orthogonal-pair interferometer.

pairs of elements.[2] The phase meter used is a twin-channel d.f. receiver suitably modified to indicate directly on the cathode-ray-tube display the phase angle within $\pm\pi$ limits. The main modification required is the addition of a sum-and-difference unit in the intermediate-frequency section to convert relative phase to relative amplitude, which can be observed directly in terms of an angle on the display. Some care is required in order to reduce instrumental errors due to different electrical lengths of coaxial cables between antenna elements and receiver, to slight differences in matching into the two receiving channels over the band, and to differences in antenna-element impedance.

Since the aperture of the array tested is about 4λ at 10 Mc/s, it is obvious that there are a number of ambiguities of bearing and elevation associated with each set of phase readings. In order to assess the error performance it is necessary to have a fairly close estimate of the bearing and, sometimes, also of the elevation of each path, so that ambiguous results can be eliminated.

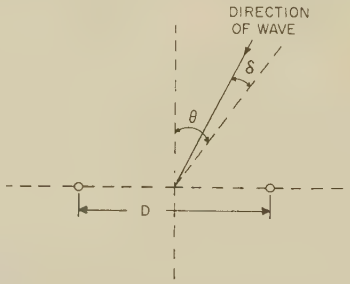


Fig. 2.—Two-element interferometer.

The phase difference in terms of phase cycles between the two elements of an interferometer (Fig. 2) is given by

$$\phi_1 = \frac{D}{\lambda} \sin\theta \cos\delta$$

where θ and δ are angles of azimuth and elevation, respectively, D is the aperture, and λ is the wavelength of the received wave. A second interferometer in the same plane but orthogonal to the first one produces a phase difference of

$$\phi_2 = \frac{D}{\lambda} \cos\theta \cos\delta$$

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Mr. McLeish and Mr. Burtnyk are in the Radio and Electrical Engineering Division of the National Research Council, Canada.

Therefore

$$\tan\theta = \frac{\phi_1}{\phi_2} \text{ and } \cos\delta = (\phi_1^2 + \phi_2^2)^{1/2}\frac{\lambda}{D}$$

The nomogram, Fig. 3, is based on these equations and is a convenient means for arriving at the approximate total phase

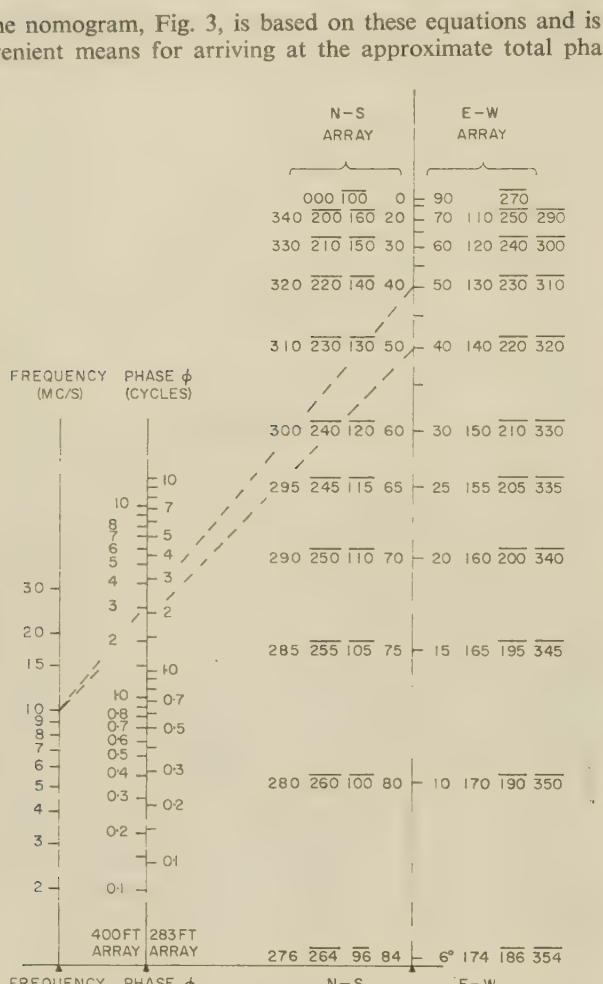


Fig. 3.—Nomogram for maximum phases across 400ft and 283ft arrays.

Bearing scale numbers with bar (e.g. $\overline{230}$) indicate negative values of ϕ; e.g. at 10Mc/s and $\theta = 230°$, $\phi_1 = -3{\cdot}1$ and $\phi_2 = -2{\cdot}6$.

across each pair when the bearing is known and the elevation is assumed to be zero. The sum of the phase-meter reading and an integer should come within a factor, $\cos\delta$, of the phase derived from the nomogram. In this way the value of the integer can usually be chosen by inspection, unless knowledge of the propagation path is very vague.

The bearing and elevation are then derived from the exact phases across each pair, using the sum of the phase-meter reading and the probable integer in each case.

(2.2) Ground-Wave Errors

In order to test properly for ground-wave error, a large number of observations on both narrow- and wide-aperture systems is required. The small amount of data shown below is indicative only of the order of errors on a rather poor d.f. site, close to several small single-storey buildings and only 650ft from overhead wires. The first series of tests was made by moving a transmitter over a narrow sector (20°) in equal increments o distance tangential to the propagation path, preferably alon roads having no overhead wiring. Changes of bearing were noted, first on an Adcock direction-finder, and on a late occasion on a 400ft interferometer of the configuration shown in Fig. 1. The deviation of observations from a straight-line best fit to the data on each run is assumed to be the total error This is made up of site, reading and instrument errors. When the interferometer indication is steady, as it is on strong ground waves, the r.m.s. reading error is about 0·5 phase degrees, and instrumental phase error on a fixed frequency is about 1·0° When converted to bearing error by the factor $\lambda/2\pi D$, these errors are small compared with the estimated site error given in Table 1.

Table 1

ADCOCK AND INTERFEROMETER SITE ERRORS

Antenna system	Number of observations	Number of sectors	Frequency Mc/s	Distance miles	R.M.S. error deg
Adcock ..	52	2	5·02	1·5–3	1·5
Interferometer	70	4	5–7·7	1·5–3	0·2

Hopkins and Horner[3] have given theoretical values of site error for various d.f. antennae. These are derived on the basis of a relatively weak interference field (less than 0·2 times the main field) generated by reradiators randomly distributed in azimuth. The relative r.m.s. errors given for narrow-aperture Adcock and vertical monopole interferometer are $1 : \lambda/(\pi D)$ which in this case gives a ratio between 6 and 10 in the frequency range tested.

Errors caused by differences of electrical length of inter connecting cables from the antennae and by differences of matching at the receiver may be calibrated out at each frequency In terms of bearing error they amounted to about 0·2° r.m.s over most of the frequency band.

(2.3) Sky-Wave Errors

The direction of arrival of sky waves is affected by the tilt of the ionization gradient at points of reflection in the ionosphere. Ross and Bramley[5,6] have found that typical F-layer tilt angles are in the range 1–2° r.m.s. and the time variation has a period from 5 to 30min. This results in lateral deviations for long-distance paths (over 1000km) of 0·5–1° r.m.s. For the E-layer, measurements show that tilts are not more than about one-third of those for the F-layer, and because the angle of elevation of the E-reflected path is also lower, the lateral deviations are very small.

The effect of wave interference on the output phase indication of the interferometer depends on the relative amplitudes of the interfering waves and on their angular distribution. To arrive at estimates of error, it has been assumed that the relative phases of interfering waves are random. The following comments are intended only to indicate some error limits.

The case where interference is weak and random in angular distribution about the main mode has been mentioned in the previous Section. When weak interference is confined to very narrow angular deviation, α, from the main mode so that α is less than $\lambda/2\pi D$, where λ is the wavelength and D is the aperture, then the indicated bearing is roughly the same as that obtained with a narrow-aperture system. The reason is that in both cases the correlation of the interference field across the aperture is high. When α is greater than $\lambda/2\pi D$, the error is reduced towards that for random angular distribution of the interfering waves.

Interference between modes of comparable amplitude cause

severe phase fluctuation in the interferometer output. Bain[4] has studied the effects of relative mode strength and azimuth and elevation angle difference upon the error of a wide-aperture phase-sensitive system. The advantage over the Adcock varies considerably with these parameters. Typical values of the ratio of r.m.s. errors for a 4λ array range from 1 to 7. In every case the mean value of the phase difference observed in the output of the interferometer is that due to the strongest mode.

In a preliminary test in July–August, 1960, observations in the 8–20 Mc/s frequency range were made on commercial transmissions, most of which originated at distances beyond 700 km. The array of Fig. 1 was used, with vertical monopole elements. Each reading was the operator's estimate of the average of the rapid fluctuations of phase after 30–50 sec of observation. A standard deviation of $1{\cdot}1°$ and a systematic error of $-0{\cdot}05°$ were obtained on 93 readings.

Although it was not possible to identify the modes by which each transmission was received, an attempt was made in a second test to separate the overall effects of lateral deviation and wave interference for observations taken during a nine-day period (21st–29th November, 1960). Reception conditions were about average during this period. Advantage was taken of the configuration of the array to obtain four phase readings, two for the orthogonal pair EW–NS, and two for the pair EN–NW, which have a smaller aperture. Again the operator was permitted to observe phase fluctuations for 30–50 sec in making a phase estimate for each reading. Because wave-interference effects are relatively rapid, the four readings are independent in this regard but lateral-deviation errors are common. A total of 211 observations were made on 95 different transmitters and the distribution of the data with respect to radio frequency is shown in Fig. 4.

Fig. 4.—Distribution with frequency.

The histograms in Fig. 5 show distributions of error with respect to great-circle bearing for (*a*) the 400 ft orthogonal pair and (*b*) the 283 ft orthogonal pair. Six observations are classed as 'wild' for both distributions and are well outside the error limits shown. Also, there are two wild observations in (*a*) group which fall well within the error limits of (*b*) group, and three wild observations in (*b*) group which are included in the (*a*) histogram. As there was no independent measure of elevation angle of arrival, no attempt has been made to assess the accuracy in this respect.

An interesting result is obtained from the distribution of difference of error between the two arrays. The correlation coefficient between the two sets of observations is

$$C = \frac{v_a + v_b - v_d}{2\sqrt{(v_a v_b)}} = 0{\cdot}51$$

where v_a, v_b and v_d are the squares of standard deviations (variances) of the two error distributions of Fig. 5 and the error difference distribution, respectively. The correlated part of each distribution (*a*) and (*b*) is

$$V_{cor} = C\sqrt{(v_a v_b)} = 0{\cdot}86\,\text{deg}^2$$

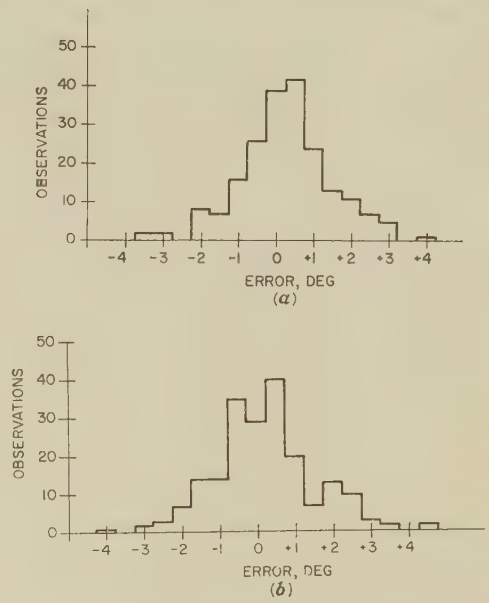


Fig. 5.—Error distribution.

(*a*) 400 ft orthogonal array.
Standard deviation = $1{\cdot}24°$.
Systematic error = $+\,0{\cdot}258°$.

(*b*) 283 ft orthogonal array.
Standard deviation = $1{\cdot}35°$.
Systematic error = $+\,0{\cdot}21°$.

and the uncorrelated parts are

$$(v_a)_{un} = 0{\cdot}68\,\text{deg}^2 \text{ and } (v_b)_{un} = 0{\cdot}97\,\text{deg}^2$$

The correlated variance may be attributed almost wholly to lateral deviation. The uncorrelated variances are due mostly to wave interference, and it is probable that the difference between them is significant, the larger interferometer having the lower fluctuation.

Further illustration of the effect of phase fluctuations on reading accuracy is contained in the histograms of Fig. 6. Here

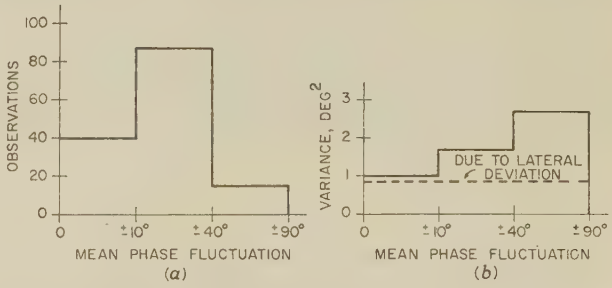


Fig. 6.—Distribution according to operator's estimate of mean phase fluctuation.

(*a*) Number of observations.
(*b*) Error variance.

the distributions of (*a*) numbers of observations and (*b*) error variance are with respect to the operator's estimate of observed mean phase fluctuation. Of the eight wild results in this group, all but one were reported by the operator to be in the class having largest fluctuations. A satisfactory phase-averaging scheme would probably reduce both the variance and the number of wild observations. This view is supported by the recent work of Bain,[7] in which the rapid variance of a set of observations was less than $0{\cdot}1\,\text{deg}^2$ when each was the mean of 70 samples taken at 3 sec intervals.

The systematic errors are large enough to be significant. Because of the rather uneven azimuth and time distributions of the data, shown in Fig. 7(*a*) and (*b*), it is quite possible that this error is due to the deviation which occurs on S–N paths in late afternoon hours caused by an E–W gradient in the ionosphere.

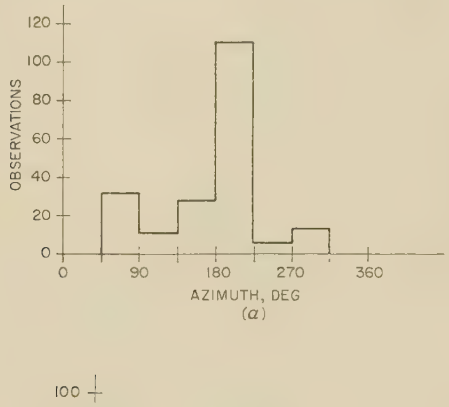


Fig. 7.—Distribution of observations.
(*a*) With azimuth.
(*b*) With time of day.

As a rough check on the effect of distance on accuracy, the observations were divided into two groups; those for transmissions from distances between 440 km (the nearest transmitter) and 1 000 km, and those for transmissions from beyond 1 000 km. The standard deviations of the two groups are 1·4° and 1·2° respectively. Although tilts produce larger bearing errors for short-distance (high-elevation-angle) paths, the longer paths are more likely to support multimode propagation and wave-interference errors will become more important. This is further illustrated in the histogram, Fig. 8, which shows only a slight

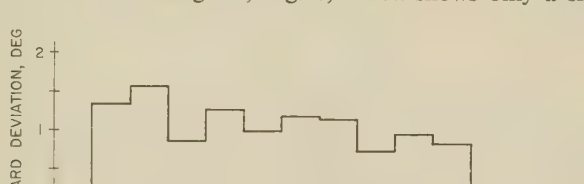


Fig. 8.—Distribution of standard deviation of error with frequency.

reduction of standard deviation of error with frequency on the 400 ft interferometer.

Using an Adcock antenna, a similar group of observations was obtained earlier at the same site. Reception conditions were much the same as during the interferometer test. The overall variance of 485 observations was 12·1 deg^2. This can be compared with the overall variance obtained on the 400 ft interferometer in the second test of 1·54 deg^2. If the lateral-deviation variance can be assumed to be about the same for the Adcock test as that obtained for the interferometer test, 0·86 deg^2, the residual variances in each case are 11·2 deg^2 and 0·68 deg^2. These may be attributed mostly to site and sky-wave interference effects since instrumental error variance is less than 10% of the residual in each case. The ratio of r.m.s. errors is about 4 : 1.

(3) CONCLUSIONS

Tests on ground-wave signals at one frequency where the aperture is about 3λ give a site error of about 0·2° r.m.s. The theoretical site-accuracy improvement of the interferometer over a narrow-aperture (Adcock) system is supported by the experimental evidence, although this is too sparse for quantitative evaluation.

On sky-wave signals the bearing errors have an r.m.s. value of 1·2° on long-distance transmissions (beyond 1 000 km). This can be compared with the results of similar accuracy studies on Adcock systems for which the mean value of standard deviation of error is about 3·5°. Shorter-distance transmissions show a distinct rise in error, probably due to ionospheric tilts.

Correlation of two groups of data on two sets of orthogonal pairs reveals a common variation, due to tilts, of about 0·9° r.m.s. This is within the range of expected values for oblique transmissions over the distances involved.

There are insufficient data to measure the effect of an aperture difference of $\sqrt{2}$: 1 on wave-interference reduction, although the larger array apparently has less of this error. Cases of wide bearings were usually associated with large phase fluctuations common to observations on both arrays.

The interferometer shows a substantial improvement in wave-interference error reduction over the narrow-aperture Adcock, but before it can be used as a direction-finder a satisfactory method of resolving the ambiguities must be found.

(4) ACKNOWLEDGMENT

The authors acknowledge many helpful discussions with their colleagues, particularly Mr. G. Evans and Mr. J. Wolfe.

(5) REFERENCES

(1) Bain, W. C.: 'The Theoretical Design of Direction-Finding Systems for High Frequencies', *Proceedings I.E.E.*, Paper No. 1960 R, January, 1956 (**103** B, p. 113).

(2) Ross, W., Bramley, E. N., and Ashwell, G. E.: 'A Phase-Comparison Method of Measuring the Direction of Arrival of Ionospheric Radio Waves', *ibid.*, Paper No. 1134 R, July, 1951 (**98**, Part III, p. 294).

(3) Hopkins, H. G., and Horner, F.: 'Direction-Finding Site Errors at Very High Frequencies', *ibid.*, Paper No. 773 R, February 1949 (**96**, Part III, p. 321).

(4) Bain, W. C.: 'The Calculation of Wave-Interference Errors on a Direction-Finder Employing Cyclical Differential Measurement of Phase', *ibid.*, Paper No. 1545 R, September, 1953 (**100**, Part III, p. 253).

(5) Bramley, E. N., and Ross, W.: 'Measurements of the Direction of Arrival of Short Radio Waves Reflected at the Ionosphere', *Proceedings of the Royal Society*, A, 1951, **207**, p. 251.

(6) Bramley, E. N.: 'Direction-Finding Studies of Large-Scale Ionospheric Irregularities', *ibid.*, 1953, **220**, p. 39.

(7) Bain, W. C.: 'Phase Difference Observations at Spaced Aerials and their Application to Direction Finding', *Journal of Research of the National Bureau of Standards*, 1961, **65** D, p. 229.

(6) APPENDIX

Ambiguity Resolution for Orthogonal Pairs

If the aperture, D, is greater than $\lambda/2$ the values of phase, ϕ_1 and ϕ_2, across orthogonal pairs cannot be determined uniquely by a single set of phase-difference measurements. The phase meter measures $\Delta\phi_1 = \phi_1 - n_1$, where n_1 is an integral number of phase cycles. The absolute value of n_1 cannot be larger than $D/\lambda - \Delta\phi$ and since it may have either sign, the total number of possible values is $2n_{max} + 1$ when no further information is available. An approximate bearing, θ_a, from a narrow-aperture Adcock system further limits the ranges of n_1 and n_2 for orthogonal pairs by setting limits to the ratio

$$\tan \theta_a \simeq \frac{n_1 + \Delta\phi_1}{n_2 + \Delta\phi_2}$$

The signs of numerator and denominator are also determined by θ_a, except when θ_a approaches 0°, 90° or their reciprocals. The approximate maximum values of n (for ground-wave propagation) are also set by θ_a and can be found from the nomogram in Fig. 3.

A rough value for angle of elevation would resolve the remaining ambiguities, since $\cos \beta = (\phi_1^2 + \phi_2^2)^{1/2}\lambda/D$. This could be obtained separately by a vertically spaced pair of elements. Perhaps both bearing and elevation can be more conveniently arrived at with sufficient accuracy using a small interferometer, not more than $\lambda/2$ in aperture, which would have no ambiguities. In practice the choice may rest on the quality of the site, which limits the accuracy of elevation-angle measurement by a horizontal-antenna array.

THEORY OF TIME-DELAY NETWORKS

By W. T. J. Atkins, B.Sc.(Eng.), Member.

(*Communication received 16th January*, 1961.)

In a paper* I described a somewhat specialized automatic oscillograph and included an Appendix briefly outlining the theory of time-delay networks of the kind employed in the instrument as accessories. I gave, as the unit-function response of an idealized time-delay network consisting of n similar sections in cascade, the expression

$$\frac{1}{\pi}[\mathrm{Si}(\omega_c t + n\pi) + \mathrm{Si}(\omega_c t - n\pi)] \quad \text{(formula B)}$$

where $\omega_c = 2\pi \times$ cut-off frequency.

In the discussion on the paper, the late Dr. S. Whitehead commented that I had departed from what he asserted was the established response formula, namely

$$\frac{1}{2} + \frac{1}{\pi}\mathrm{Si}(\omega_c t - t_0) \quad \text{(formula A)}$$

where $t_0 =$ the delay time. This formula was quoted from Professor E. A. Guillemin's 'Communication Networks', Vol. 2, where it appears in eqn. (984*b*).

I debated the relative validity of the two expressions with Dr. Whitehead, as was sufficiently reported at the time in the *Proceedings*, and also had a fairly lengthy private correspondence with him on the subject, but we were unable to reach satisfactory agreement and so allowed the question to drop, since neither of us could then spend more time on it.

Having lately had the opportunity to look again into this matter, and to consult further references, I have seen Guillemin's more recent book† and find that he there gives as the impulse response of a network having exactly the same idealized specification as mine a new formula [Ch. 15, eqn. (154)]:

$$\frac{\omega_c}{\pi}\left[\frac{\sin \omega_c(t - t_0)}{\omega_c(t - t_0)} + \frac{\sin \omega_c(t + t_0)}{\omega_c(t + t_0)}\right]$$

This is effectively identical with formula B, bearing in mind that $t_0 = n\pi/\omega_c$ and that impulse response is the differential of unit-function response. It is unfortunate that Guillemin does not explicitly relate these two versions to each other, though a sufficiently perspicacious reader will gather that the earlier one is an approximation to the later. From a practical point of view, the matter is only of significance when the number of network sections is small, though then the superior accuracy of the more complex form will be manifest to anyone having the patience—or the computing facilities—to work out the response of actual networks of one or two sections.

It is nevertheless of interest to find the theoretical explanation for this puzzling discrepancy between two allegedly correct answers. When shorn of mathematical technicalities it can be stated as follows. Formula B (calculated by Laplace transform of Carson integral) refers to a network comprising a finite number of sections and correctly terminated by a resistor matched to the characteristic impedance of the network. Strictly, the network components should be pure reactances but may, without significant error, have balanced losses conforming to the distortionless condition. Formula A (calculated by Dirichlet integral) refers to an infinitely extended network of purely reactive components producing no losses anywhere. It is implied that time extends from infinite past to infinite future, and that the unit-function signal progresses through it without attenuation, becoming available for study at the instant of zero. This, in turn, imposes a symmetry on the network by requiring the addition of a mirror-image of it, extending backwards to infinity from its input terminals, within which all the signals it is to carry have been approaching since infinite past. Consequently, by the time the unit-function signal reaches the moment of zero it has already acquired its full measure of distortion, and it is therefore preceded as well as followed by a system of ripples symmetrical about the mid-point.

It should be remembered that Guillemin's earlier work is mainly concerned with the spectral response of filter networks to cisoidal excitation over a band of frequencies, and his signals usually form the modulation envelope of a carrier. They enter the network under examination via some other channel of roughly similar frequency response, and hence there is no error of practical consequence in assuming a symmetrical build-up and die-away of transients. The requirements of networks designed specifically for faithful reproduction of the transients themselves are naturally more stringent, and where Guillemin tackles that problem, as he does towards the end of 'Synthesis of Passive Networks', he takes proper account of these matters, and indeed his elegant system of design for ripple control in time-delay networks makes the methods of my 1948 paper appear crude in the extreme.

It will be concluded that formulae A and B are both right, but, as sometimes happens in other than mathematical circles, one is more right than the other.

* Atkins, W. T. J.: 'An Oscillograph for the Automatic Recording of Disturbances on Electric Supply Systems', *Proceedings I.E.E.*, Paper No. 787 M, December, 1948 (96, Part II, p. 276).

† Guillemin, E. A.: 'Synthesis of Passive Networks' (Chapman and Hall, 1957).

621.387.322

The Institution of Electrical Engineers
Paper No. 3420 E
Jan. 1961

COMPARISON OF ARGON, KRYPTON AND XENON AS ADMIXTURES IN NEON GLOW-DISCHARGE REFERENCE TUBES

By F. A. BENSON, D.Eng., Ph.D., Associate Member, and G. P. BURDETT, B.Eng., Student.

(*The paper was first received* 11*th March, in revised form* 29*th July, and in final form* 14*th October*, 1960. *It was published in January*, 1961, *and was read before the* Electronics and Communications Section 8*th May*, 1961.)

SUMMARY

Special glow-discharge reference tubes containing neon-krypton, neon-xenon, neon-0·3% argon-krypton and neon-0·3% argon-xenon mixtures have been examined, the krypton or xenon content varying from 0·001 to 1%. The tubes had molybdenum electrodes and the total gas pressure was either 40 or 20 mm Hg. The striking voltage, running-voltage/current, running-voltage/temperature, initial drift, impedance/frequency and noise-voltage/current characteristics have been determined. The results of the studies are presented and discussed and are compared with those previously obtained for neon-argon tubes. It is concluded that certain desired tube characteristics can frequently be obtained by the correct choice of the gas mixture.

(1) INTRODUCTION

The characteristics and limitations of glow-discharge tubes have been studied in some detail by several independent investigators.[1–8] Early work indicated that certain parameters, e.g. the gas filling, gas pressure and cathode material, greatly affected tube characteristics. It also suggested that the argon content of the gas filling was another very important parameter in determining the characteristics of a tube. As a result, measurements were made on special neon-argon- and helium-argon-filled tubes of the stabilizer type with cerium cathodes in which only the argon content was varied.[9] A fuller study of the effects of varying the argon content in neon-argon tubes was carried out using special high-stability reference tubes with molybdenum electrodes.[10] The argon content was varied over the range 0·001–10% and it was concluded that its optimum value, taking all the characteristics into account, was of the order of 1%. Other information on the properties of tubes containing neon-argon and helium-neon mixtures can be found elsewhere.[6, 11, 12]

Studies similar to those by Benson and Chalmers[10] have now been made using neon-krypton, neon-xenon, neon-0·3% argon-krypton and neon-0·3% argon-xenon mixtures. With neon-argon fillings it had been shown previously[10] that the most efficient discharge, i.e. minimum running voltage, was obtained with an argon content of 0·3%. The particular triple mixtures used during the present investigations were chosen, therefore, to determine if the additions of krypton or xenon to neon-0·3% argon tubes produced the same effects as when they were included in pure-neon fillings.

The tubes used were of the reference type in which a sputtered layer of molybdenum covers the glass walls.[1] The molybdenum electrodes consisted of a cylindrical cathode 7·5 mm in diameter and 10 mm long and a rod anode 1 mm in diameter mounted concentrically. The volume of the glass envelope was about 8 cm^3. A mercury diffusion pump was used during the manufacture of the tubes, so an adequate cold trap was fitted to avoid mercury-vapour contamination. Four tubes of each of the following gas fillings and with a total gas pressure of 40 mm Hg were examined:

(*a*) Pure neon, and neon plus 0·001, 0·01, 0·1 and 1% krypton or xenon.

(*b*) As in (*a*) with a neon-0·3% argon mixture instead of pure neon.

The striking voltage, running-voltage/current, running-voltage/temperature, initial drift, impedance/frequency and noise-voltage/current characteristics were measured for all the tubes.

In addition, it was decided to study the effect of a pressure variation and so four tubes of each of the following gas fillings were examined—neon plus 1, 3 and 10% of argon or krypton or xenon. These tubes were exactly the same as those above except that the total gas pressure was 20 mm Hg. The running voltage and its change with temperature were measured for each tube.

The results of the work are presented and discussed and are compared with those previously obtained for neon-argon tubes.[10]

(2) MEASUREMENTS

In the present tubes the pre-breakdown gaps* mentioned by Benson and Chalmers[10] were controlled at 1 mm so that striking voltages were recorded. These measurements were repeated several times for each tube. The effect of light on striking voltage was very small, except for one or two tubes with low percentages of admixture which showed a much higher striking voltage (of the order of 50 volts) in total darkness. Because of this, the spread in results in total darkness was much greater than in normal daylight, and for this reason the results in daylight are given.

The running-voltage/current characteristics, the variations in running voltages with temperature and the initial drifts of the tubes were determined using measuring techniques described in previous papers.[13, 14]

The initial drift and effect of temperature on the running voltage were measured at 4 mA, this current being chosen because it was always in the normal-glow region.

The effective a.c. resistances and reactances of the tubes were measured over the frequency range 5 c/s–30 kc/s. For the range 300 c/s–30 kc/s the modified form of Owen bridge network described by Benson and Chalmers[10] was used. The tubes were operated at a mean current of 3 mA so that the results could be compared with those previously obtained for neon-argon mixtures, and the superimposed alternating current was kept at 0·3 mA, this value having been chosen as a compromise between an easily detectable null point and minimum distortion. When calibrated against standard resistors and inductors this bridge circuit gave a maximum error of 2%. Measurements with it were limited at high frequencies by the detector sensitivity and the effect of stray capacitance, and at low frequencies by increased distortion and inaccuracies due to the bridge arms

Dr. Benson and Mr. Burdett are in the Department of Electrical Engineering, University of Sheffield.

* The pre-breakdown gap, i.e. the distance between the anode strapping and the cathode, has been illustrated elsewhere.[26]

having widely differing impedances. The components of the tube impedances were determined over the frequency range 5c/s–2kc/s using a commercial phase-meter which incorporated a tunable filter. The results obtained by the two methods in the range 300c/s–2kc/s showed close agreement.

Noise measurements were made with a calibrated amplifier-detector unit having an equivalent noise bandwidth of 87kc/s. The noise voltages were recorded for different tube currents over the range 0·5–5mA, for a fixed noise load resistance of 37·5 kilohms.

(3) RESULTS

Striking voltages, minimum running voltages, changes in running voltage with temperature, impedances at 30kc/s, and mean-noise voltages at 3mA are given in Table 1 for neon-xenon, neon-krypton, neon-argon-xenon and neon-argon-krypton tubes at 40mm Hg. Similar results for neon-argon, except for striking voltage which was not recorded, taken from Reference 10, are also included for comparison. The spread in results for tubes of the same type is indicated. Table 2 shows the minimum running voltages and changes in running voltage with temperature for neon-xenon, neon-krypton and neon-argon tubes at 20mm Hg. The spread in results is also indicated.

Running-voltage/current and change in running-voltage/temperature characteristics are illustrated in Figs. 1 and 2, respectively, for neon-krypton tubes. Neon-xenon results are similar and are therefore not shown. These curves are for a typical tube chosen from a batch of four, and although there is a spread in results, the forms of the curves do not differ appreciably for tubes of the same batch. Each curve has been drawn through all the experimental points, and during the measurements the current was varied gradually so that any voltage jumps could be detected.

Figs. 3 and 4 show impedance results for neon-krypton and neon-xenon tubes, respectively, at higher frequencies, and the low-frequency results are shown in Fig. 5. Examples of noise-voltage/current curves are shown in Fig. 6 for neon-krypton and neon-argon-krypton tubes.

(4) DISCUSSION OF RESULTS

(4.1) Striking and Running Voltages

Small additions of krypton or xenon in a pure-neon glow-discharge tube lower the striking and running voltages (called the Penning effect[5]) in a manner similar to argon. The first Townsend ionization coefficient, α, may be divided into α_e, the ionization coefficient for direct ionization by electrons, and α_m, the ionization coefficient for ionization of admixture atoms by metastable atoms. In the pure gas, α_m will be zero but, as the amount of admixture is increased, α_m, and therefore α, will increase, provided that the ionization potential of the admixture atoms is smaller than the potential of the neon metastable atoms. α will reach a maximum when there are sufficient admixture atoms in the cathode-fall region, where most of the ionization occurs, to employ all the metastable atoms. The concentration of admixture atoms around the cathode, called cataphoresis,[5] ensures that this occurs at a lower percentage of admixture than would otherwise be the case. The effect of the presence of the admixture on the striking voltage is much greater than on the running voltage. The minimum striking voltages were obtained with an admixture content of the order of 0·1%, the minimum occurring at a lower percentage of xenon than krypton.

In the case of running voltage the Penning effect was very noticeable for argon (Table 1), less marked for krypton and very small for xenon. The excitation potential of neon metastable atoms is 16·53 volts, whereas the ionization poten-

Table 1

CHARACTERISTICS OF NEON-XENON, NEON-KRYPTON, NEON-ARGON, NEON-ARGON-XENON and NEON-ARGON-KRYPTON TUBES WITH A TOTAL PRESSURE OF 40 MM HG

Gas	Striking voltage	Minimum running voltage	Change of running voltage with temperature	Impedance at 30kc/s	Mean noise voltage
	volts	volts	mV per deg C	kilohms	μV
Pure neon	220 to 234	107·1 to 107·8	−17·1 to −17·5	6·01 to 6·50	234 to 253
Neon plus					
0·001% xenon	194 to 200	104·1 to 106·9	−19·5 to −20·2	6·11 to 7·75	220 to 254
0·01% xenon	128 to 136	98·75 to 101·05	−16·5 to −31·75	10·3 to 11·0	212 to 230
0·1% xenon	136 to 138	96·7 to 97·6	−2·25 to −5·75	5·77 to 6·55	282 to 297
1% xenon	150 to 155	99·1 to 100·1	+7·5 to +13·0	1·95 to 2·02	344 to 361
Neon plus					
0·001% krypton	191 to 200	103·90 to 104·55	−18·0 to −20·5	6·06 to 6·50	239 to 250
0·01% krypton	142 to 151	100·35 to 102·80	−18·5 to −20·0	6·25 to 9·80	209 to 248
0·1% krypton	121 to 127	88·25 to 91·55	−6·0 to −12·0	6·90 to 10·64	203 to 229
1% krypton	138 to 162	92·95 to 97·80	−0·5 to + 0·5	2·05 to 2·35	298 to 306
Neon plus					
0·001% argon		103·2 to 104·6	−20·0 to −20·8	6·24 to 6·35	234 to 237
0·01% argon		99·3 to 100·5	−22·0 to −24·6	9·00 to 9·35	206 to 208
0·1% argon		84·4 to 85·6	−10·2 to −13·3	8·30 to 11·05	178 to 192
1% argon		83·7 to 85·3	−1·07 to −2·17	2·19 to 2·35	251 to 262
Neon-0·3% argon	108 to 115	81·7 to 83·3	−9·0 to −10·2	3·56 to 4·06	210 to 264
Neon-0·3% argon plus					
0·001% xenon	120 to 121	84·6 to 84·8	−8·0 to −16·0	7·5 to 8·05	210 to 215
0·01% xenon	114 to 120	84·40 to 84·65	−6·0 to −10·0	7·38 to 7·46	208 to 212
0·1% xenon	134 to 141	93·2 to 94·5	+7·0 to +12·0	3·0 to 3·55	295 to 322
1% xenon	140 to 145	98·0 to 101·9	+9·0 to +12·0	2·25 to 2·30	340 to 357
Neon-0·3% argon plus					
0·001% krypton	118 to 122	84·60 to 85·75	−11·5 to −13·5	8·3 to 9·0	200 to 243
0·01% krypton	112 to 122	84·95 to 85·80	−9·5 to −13·0	8·2 to 8·95	200 to 238
0·1% krypton	114 to 124	85·1 to 85·8	−5·0 to −6·0	4·8 to 5·2	230 to 244
1% krypton	136 to 158	90·35 to 93·40	−0·07 to −0·35	2·05 to 2·25	294 to 300

The change of running voltage with temperature was calculated over the range 20–50° C.

tials of argon, krypton and xenon are 15·8, 14·0 and 12·1 volts respectively.[16] It appears at first sight, therefore, that the effect should be greatest with xenon since there is the largest difference in this case between the ionization potential and the excitation potential of the neon metastable atoms. In fact, a resonance condition prevails. The potential of the neon metastable atoms is only slightly higher than the ionization potential of the argon atoms so that the probability of ionization is high. Massey and Burhop[17] show that the ionization cross-sections for a particular process of charge or excitation transfer fall off rapidly for similar systems as the change in internal energy increases, exhibiting a maximum at or near exact energy resonance.

The ionization of argon atoms by neon metastable atoms is quoted as being close to exact resonance since the change in internal energy is very small. It has also been shown theoretically[18] that, for small amounts of argon in neon (less than 0·01%), the probability that a neon metastable atom will ionize an atom of the admixture is very close to unity. This probability will be less for krypton and very small for xenon, from a consideration of the changes in internal energy involved.

For a particular percentage of admixture the minimum running voltage for the neon-krypton tubes was slightly greater than for the neon-argon ones in all cases, because of the smaller Penning effect in neon-krypton. Similarly, the minimum running voltage for the neon-xenon tubes was higher than that for a corresponding percentage in the neon-krypton ones, except for 0·01% xenon when it was slightly lower. This unusual behaviour at 0·01% xenon also occurs in the other characteristics.

(4.2) Running-Voltage/Current Curves and Initial Drifts

The range of current for the normal-glow discharge and the full-glow current, which is a measure of the current density, depend greatly on the percentage of admixture. The current at which the glow jumped to the outside surface of the cathode cylinder was measured for each tube and was found to be greatest for tubes containing 0·001% admixture; it then decreased to a minimum at about 0·1% and finally began to increase again for higher percentages. This also applied to the current at which the running voltage was a minimum, the minimum value again occurring for an admixture content around 0·1%. The actual values of the currents at which these two phenomena occur also depend on the nature of the admixture, but there was a tendency for the currents to be highest for tubes containing xenon and lowest for those containing argon.

The movement of the glow from the inside to the outside of the cathode is accompanied by a sudden decrease in running voltage due to a decrease in current density.[10] This is not shown on all the curves in Fig. 1. Prior to this, the running voltage rises with increasing current, owing to the edge effect at the cathode. A further step was observed in the characteristics for pure-neon tubes and tubes containing 0·001 and 1% of both krypton and xenon. The current densities in these tubes were comparatively high, and thus sputtering was more difficult to carry out and the cathode surfaces may not have been perfectly clean. This conclusion is supported by the results reported for neon-argon tubes with cerium cathodes[9] where steps were found in the characteristics and are now thought to be due to unevenly-prepared cathode surfaces, for which the sputtering and preparation procedures were influenced by the argon content to differing extents.

Hysteresis effects were observed to be slightly smaller for neon-krypton and neon-xenon tubes than for argon ones. The maximum difference in the running voltage, at a given current, for increasing and decreasing currents was less than 0·1 volt. For all admixtures over the current range 2–10 mA, the regulation was least for tubes containing 1% admixture.

It is interesting to examine the spread between tubes of the same type. When a small amount of admixture is present, tubes show a fairly large spread, probably due to slight variations in the amount of admixture. Tubes containing 0·1% admixture, particularly neon-xenon ones, show a much smaller spread. Neon-1% argon and neon-1% xenon tubes also display a small spread, and the minimum spread therefore coincides approximately with the minimum in the running-voltage/admixture-content curve and hence with the region where the discharge is most efficient. There is a very large spread in the results for

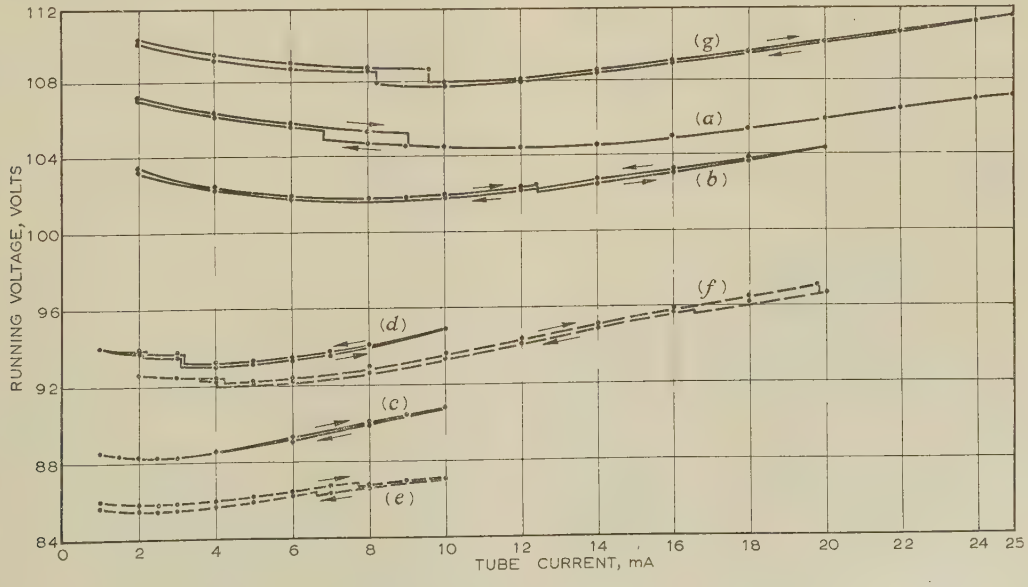


Fig. 1.—Running-voltage/current characteristics.

(*a*) Neon-0·001% krypton.
(*b*) Neon-0·01% krypton.
(*c*) Neon-0·1% krypton.
(*d*) Neon-1% krypton.
(*e*) Neon-0·3% argon with 0·001, 0·01 and 0·1% krypton.
(*f*) Neon-0·3% argon-1% krypton.
(*g*) Pure neon.

tubes containing 1% krypton, probably because an anode fall is just beginning to form, whereas in neon-argon tubes this has not started and in neon-xenon tubes it is complete.

The initial drift of running voltage was found to be small in magnitude (0·6 volt maximum) and short in duration (2–3 min) for all tubes. The drift was smallest for all three admixtures when the content was between 0·1 and 1%; for this range the drift was less than 0·1 volt and was complete after about 1 min.

(4.3) Running-Voltage/Temperature Curves

It has been suggested by Jurriaanse[19] that all tubes should have a negative temperature coefficient. Most of the voltage drop in a glow-discharge tube occurs in the cathode-fall region and depends on the gas density there. When the tube-envelope temperature increases, the gas density near the cathode will increase. Jurriaanse calculated the temperature coefficient of running voltage by considering this density change in the cathode region and relating it to the slope of the running-voltage/pressure curve at the initial tube pressure. Benson and Chalmers[10] thought that, since it is necessary to use running-voltage/pressure characteristics to determine temperature coefficients, it was better first to estimate the change in gas pressure, and not density, with varying tube-envelope temperature. A modified

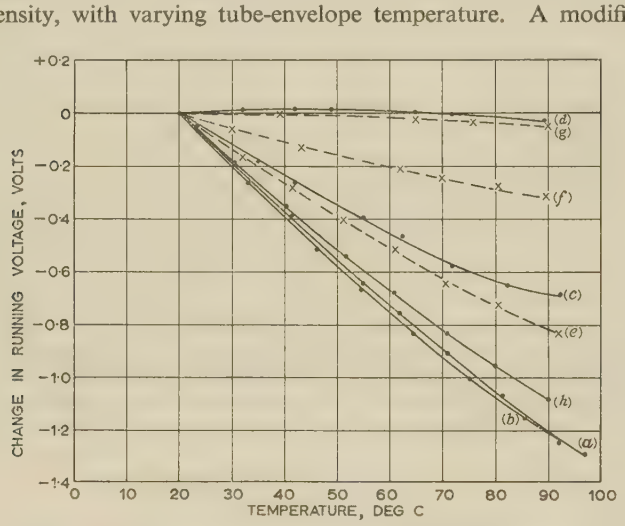


Fig. 2.—Running-voltage/temperature characteristics.
(*a*) Neon-0·001% krypton.
(*b*) Neon-0·01% krypton.
(*c*) Neon-0·1% krypton.
(*d*) Neon-1% krypton.
(*e*) Neon-0·3% argon with 0·001 and 0·01% krypton.
(*f*) Neon-0·3% argon-0·1% krypton.
(*g*) Neon-0·3% argon-1% krypton.
(*h*) Pure neon.

form of Jurriaanse's theory was therefore developed,[10] from which the decrease in voltage/temperature coefficient with increasing argon content in neon-argon tubes could be explained. The theory shows that the temperature coefficient of a tube should be proportional to the slope of its running-voltage/pressure curve. It has been pointed out by Smith[20] that density changes and not pressure changes are responsible for the variation of running voltage with temperature. This is possibly true, but for the small changes in temperature around the cathode, caused by the ambient-temperature variations considered, it seems justifiable to regard changes in pressure. In the present investigations, however, certain tubes showed positive temperature coefficients which were both reproducible and reversible if the envelope temperature was decreased to room temperature. In the past,[10] positive temperature coefficients have been ascribed to the release of impurities.

It was decided to examine the behaviour of the neon-1% xenon tubes at low temperature because these showed a definite positive temperature coefficient in the range 20–90° C. The tubes were plunged into a mixture of solid carbon dioxide and acetone, whereupon the running voltages dropped by about 8 volts. This may be explained as follows.

With the anode-cathode spacing and gas pressure used, the tubes are operating without a positive column, i.e. with the anode in the Faraday dark space. If the pressure is increased the discharge phenomena are compressed to the cathode.[21] At a particular pressure the anode enters the positive column and an anode fall ensues. The voltage then rises until the anode fall is complete and then the running-voltage/pressure curve continues its original slope as the pressure is further increased. The pressure at which this takes place depends on the gas filling, the addition of argon, krypton or xenon being equivalent in this respect to increasing the pressure. The anode fall occurs at a fairly high percentage of argon (greater than 1%), at approximately 1% krypton and at less than 1% xenon. Thus the neon-1% xenon tubes have an anode fall at room temperature, but if the temperature is decreased (i.e. if pressure is decreased) the anode fall will disappear and the voltage will fall.

The shape of the running-voltage/pressure curve just described has been verified for helium.[11] Work at present in progress is showing that argon gives similar results, and there is also a tendency for the effect to occur in neon-argon mixtures containing a fairly high percentage of argon.

Table 2

MINIMUM RUNNING VOLTAGE AND CHANGE IN RUNNING VOLTAGE PER DEG C OVER THE TEMPERATURE RANGE 20–50° C FOR NEON-XENON, NEON-KRYPTON AND NEON-ARGON TUBES WITH A TOTAL PRESSURE OF 20 MM HG

Gas	Minimum running voltage	Change in running voltage with temperature
	volts	mV per deg C
Neon plus		
1% xenon ..	99·0 to 99·7	0 to +1·3
3% xenon ..	104·4 to 105·1	+10·0 to +17·2
10% xenon ..	117·15 to 117·9	+1·0 to + 2·5
Neon plus		
1% krypton ..	88·95 to 90·15	−6·1 to −7·0
3% krypton ..	92·15 to 92·9	−5·5 to −7·0
10% krypton ..	99·2 to 100·65	+5·2 to +8·0
Neon plus		
1% argon ..	87·25 to 87·4	−10·5 to −18·5
3% argon ..	88·15 to 89·1	−2·75 to −4·05
10% argon ..	92·3 to 93·0	−3·3 to −4·1

The results for the tubes at 20 mm Hg (Table 2) add support to the above explanation. The neon-argon tubes showed no sign of an anode fall, even when the argon content was as high as 10%. It was also possible to have as much as 3% krypton present without having an anode fall, although with the neon-10% krypton tubes one was beginning to occur, as is shown by the positive temperature coefficient. The presence of 1% xenon in the tubes at 20 mm Hg resulted in an almost zero temperature coefficient, i.e. an anode fall was not present, but the neon-3% xenon tubes had a high positive temperature coefficient. This is because these tubes were working on the part of the running-voltage/pressure curve which has a high positive slope, corresponding to the formation of an anode fall. The neon-10% xenon tubes had a small positive temperature coefficient because the anode fall was almost complete.

Table 1 shows that admixtures of argon and krypton have

about the same effect on temperature coefficient of running voltage. The temperature coefficient for the neon-krypton tubes was found to be slightly more positive than for the neon-argon ones and consequently has a zero value at a somewhat smaller percentage of admixture. The effect of introducing xenon was again different. The temperature coefficient was more positive, except for tubes containing 0·01% of xenon which showed the largest running-voltage changes of all.

(4.4) Impedance/Frequency Characteristics

Figs. 3 and 4 show that both the resistive and reactive components, as well as the total impedance, depend on the percentage

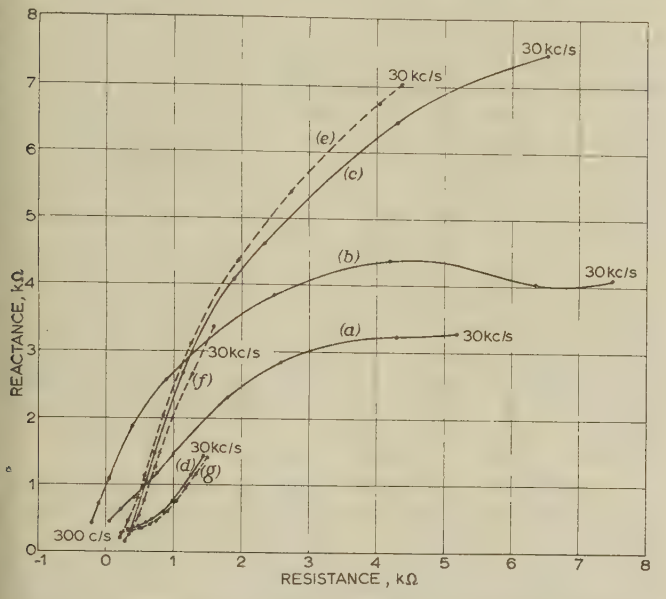


Fig. 3.—Impedance loci for tubes containing different percentages of krypton.

(*a*) Neon-0·001% krypton.
(*b*) Neon-0·01% krypton.
(*c*) Neon-0·1% krypton.
(*d*) Neon-1% krypton.
(*e*) Neon-0·3% argon with 0·001 and 0·01% krypton.
(*f*) Neon-0·3% argon-0·1% krypton.
(*g*) Neon-0·3% argon-1% krypton.

Points refer to readings at the following frequencies: 300 c/s, 600 c/s, 1 kc/s, 2 kc/s, 3 kc/s, 6 kc/s, 10 kc/s, 20 kc/s, 30 kc/s.

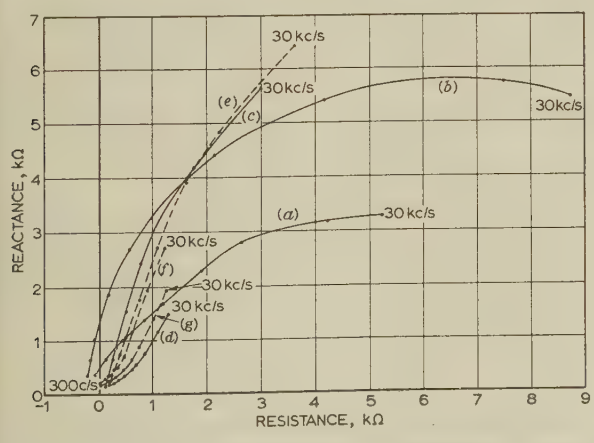


Fig. 4.—As in Fig. 3, but with xenon substituted for krypton.

of admixture. An admixture addition of 0·001% has a very small effect on the impedance, the effect being approximately the same for argon, krypton and xenon. As for other characteristics, the addition of 0·01% xenon has a different effect from the addition of 0·01% of argon or krypton. The difference is most noticeable in the reactance, which is higher than that for corresponding neon-argon and neon-krypton tubes, at a particular frequency. Further addition of xenon causes a sharp decrease in resistance and a more gradual fall in reactance. Further additions of argon and krypton cause considerable increases in reactance but gradual reductions in resistance. When the admixture percentage reaches 1% the effect of the three gases is again very similar. Table 1 shows that the addition of argon or krypton has a very similar effect on the impedance, which has a maximum at about 0·1% of admixture. The maximum impedance for the neon-xenon tubes, however, occurs at about 0·01% xenon. The low-frequency reactance/resistance curves (Fig. 5) show that as the frequency is decreased the reactance falls, until at 5 c/s the impedance of a tube is almost purely resistive.

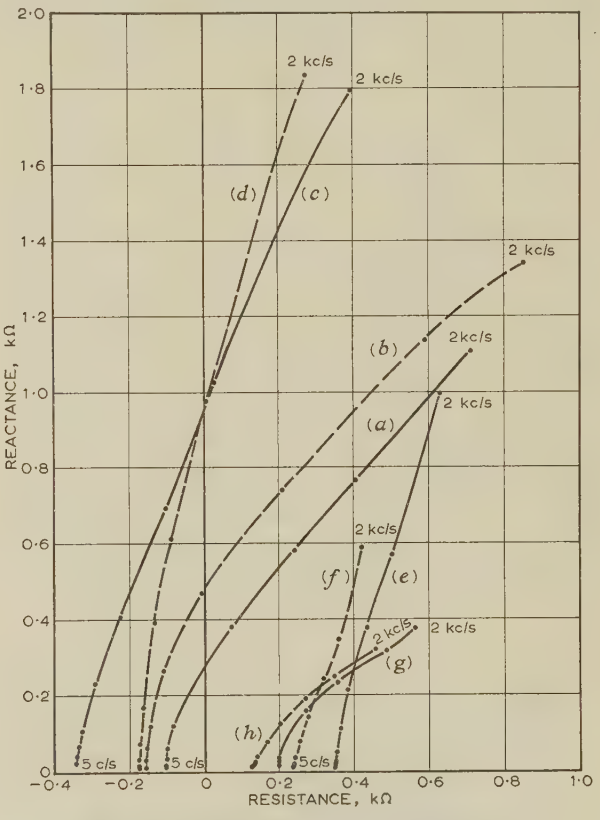


Fig. 5.—Impedance loci for low frequencies.

(*a*) Neon-0·001% krypton.
(*b*) Neon-0·001% xenon.
(*c*) Neon-0·01% krypton.
(*d*) Neon-0·01% xenon.
(*e*) Neon-0·1% krypton.
(*f*) Neon-0·1% xenon.
(*g*) Neon-1·0% krypton.
(*h*) Neon-1·0% xenon.

Points refer to readings at the following frequencies: 5 c/s, 15 c/s, 30 c/s, 60 c/s, 150 c/s, 300 c/s, 600 c/s, 1 kc/s, 2 kc/s.

A theory has been developed[22–25] to explain impedance/frequency characteristics and this predicts that the impedance, Z, has the following form if there are n delayed effects:

$$Z = a_0 - \frac{a_1}{1 + j\omega\tau_1} - \sum_{k=2}^{k=n+1} \frac{a_k}{1 + j\omega\tau_k}$$

where a_0, a_1 . . . a_k have the dimensions of resistance and $\tau_1, \tau_2 \ldots \tau_k$ are the time-constants of the various delay processes.

The second term in this expression can be regarded as the

contribution from the discharge process itself and the last term can be attributed to the delayed-effect processes. Thus, the impedance loci can consist of any number of semicircle components depending on the number of delayed effects.

It has been concluded[25] from measurements of impedance characteristics of tubes with various neon-argon gas fillings over the frequency range 300 c/s–5 Mc/s that the number of delayed effects present depends on the amount of admixture. Over the frequency range employed in the present investigations no difference could be detected in the forms of the reactance/resistance curves for corresponding percentages of argon, krypton and xenon. The differences in actual impedance values, however, suggest that the effect of the various secondary processes depends on the nature of the admixture.

(4.5) Noise Characteristics

The mean noise-voltage/current curves, examples of which are shown in Fig. 6, are similar for all three admixtures to

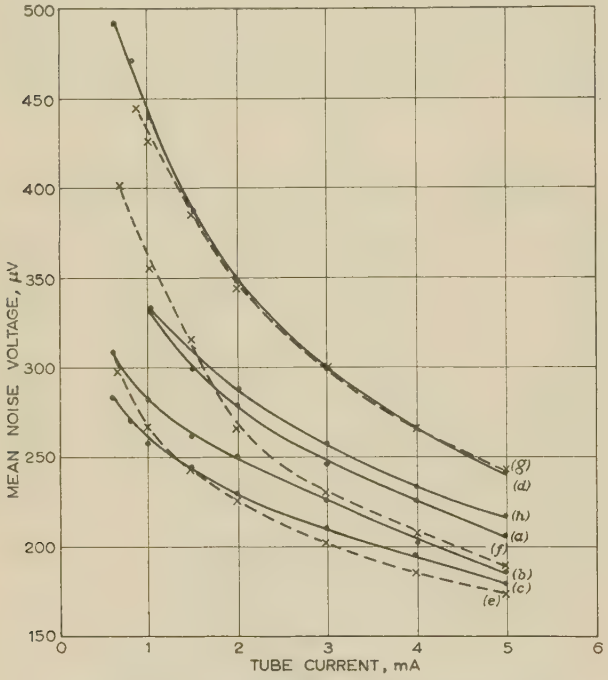


Fig. 6.—Mean noise-voltage/current characteristics.

(*a*) Neon-0·001% krypton.
(*b*) Neon-0·01% krypton.
(*c*) Neon-0·1% krypton.
(*d*) Neon-1·0% krypton.
(*e*) Neon-0·3% argon with 0·001 and 0·01% krypton.
(*f*) Neon-0·3% argon-0·1% krypton.
(*g*) Neon-0·3% argon-1% krypton.
(*h*) Pure neon.

those previously described.[10] The noise voltage increases for decreasing tube current, quite rapidly at low currents. It has been pointed out[9] that noise is a function of running voltage since both are functions of electron energy. Thus, minimum noise and minimum running voltage should occur at about the same percentage of admixture. Table 1 shows that this is true for neon-argon and neon-krypton tubes, but for the neon-xenon tubes minimum noise occurs at a lower percentage of admixture than that which gives minimum running voltage. This is difficult to explain but may have some connection with the formation of an anode fall.

Certain tubes, particularly the pure-neon ones and those containing very small amounts of admixture, exhibited internal oscillations at low currents. As mentioned already, however, these tubes were difficult to sputter and so the cathode surfaces may not have been perfectly clean. Associated with this it was observed that in these tubes the glow covered only a small area of the cathode at low currents and hence it tended to jump from one part of the cathode surface to another.

(4.6) Triple Gas Mixtures

All the characteristics show that the effect of adding krypton or xenon to the neon-argon tubes is negligible until the percentage of the second admixture is comparable with the 0·3% of argon. Tubes containing 1% of krypton or xenon behave almost as though no argon were present.

The tubes containing 0·001 or 0·01% of krypton and xenon were observed to have approximately the same running voltage but this was slightly higher than the voltage corresponding to neon-0·3% argon. The discrepancy is probably due to a slight variation in the percentage of argon present in the tubes, since it occurs throughout the characteristics. The effect of adding 0·1% of the second admixture is to produce a running voltage lying between the values for neon-0·3% argon and neon-0·1% krypton or xenon respectively. Additions of 0·1% admixture have only slight effect in the case of krypton but have a large effect with xenon. The inclusion of 0·3% argon causes small reductions in the running voltages of tubes containing 1% krypton or 1% xenon, as would be expected from the results of Table 1. It would appear that the effect on tube striking voltage of using triple gas mixtures is similar to the effect on running voltage, but there are no corresponding striking-voltage characteristics for neon-argon mixtures in the previous paper[10] for comparison with the present results.

The temperature coefficient of running voltage for the neon-0·3% argon tubes containing 0·1% krypton or xenon is approximately the same as the sum of the temperature coefficients due to the two admixtures acting independently. This behaviour does not apply to impedance or noise voltage, where the combined effect was much greater than the sum of the two individual effects.

(5) CONCLUSIONS

It has been shown that krypton and xenon can be used as admixtures in neon glow-discharge reference tubes instead of argon. One particular advantage of using the alternative gases is that tubes can be manufactured which have different running voltages but which otherwise have very similar characteristics. For example, a tube containing neon-0·1% xenon has characteristics almost identical with those of a tube containing neon-0·3% argon, except for a small increase in noise voltage which is not very important, and a much larger minimum running voltage, i.e. 97·3 volts compared with 83·7 volts. The spread in results for tubes containing neon-0·1% xenon is comparable with that for neon-0·3% argon tubes.

If small amounts of argon and krypton or argon and xenon are added to a pure-neon tube, the tube behaves as if only one admixture were present, except when the percentages of the two are comparable, when certain desirable characteristics are obtainable, e.g. low impedance.

(6) ACKNOWLEDGMENTS

The work recorded has been carried out in the Department of Electrical Engineering at the University of Sheffield. The authors wish to thank Professor A. L. Cullen for facilities afforded in the laboratories of this Department, and for his constant encouragement and advice; also Mr. P. M. Chalmers for many valuable suggestions. They also wish to acknowledge the kindness of the English Electric Valve Co., Ltd., in supplying

all the special tubes for examination, for financial assistance with the work, for the loan of equipment, and for many helpful discussions.

(7) REFERENCES

(1) BENSON, F. A.: 'Voltage Stabilized Supplies' (MacDonald, 1957).
(2) TITTERTON, E. W.: 'Some Characteristics of Certain Voltage-Regulator Tubes', *Journal of Scientific Instruments*, 1949, **26**, p. 33.
(3) BENSON, F. A., and BENTAL, L. J.: 'Glow-Discharge Stabilizers', *Wireless Engineer*, 1955, **32**, p. 330.
(4) BENSON, F. A., and BENTAL, L. J.: 'Glow-Discharge Tubes', *ibid.*, 1956, **33**, p. 33.
(5) DRUYVESTEYN, M. J., and PENNING, F. M.: 'The Mechanism of Electrical Discharges in Gases of Low Pressure', *Reviews of Modern Physics*, 1940, **12**, p. 87.
(6) KRUITOF, A. A., and PENNING, F. M.: 'Townsend Ionization Coefficient α for Neon-Argon Mixtures', *Physica*, 1937, **4**, p. 430.
(7) PENNING, F. M.: 'On the Normal Cathode Fall in Neon', *ibid.*, 1949, **15**, p. 721.
(8) JURRIAANSE, T., PENNING, F. M., and MOUBIS, J. H. A.: 'The Normal Cathode Fall for Molybdenum and Zirconium in the Rare Gases', *Philips Research Reports*, 1946, **1**, p. 225.
(9) BENSON, F. A., and GILLESPIE, E. F. F.: 'Influence of Argon Content on the Characteristics of Glow-Discharge Tubes', *Proceedings I.E.E.*, Paper No. 2394 R, September, 1957 (**104** B, p. 498).
(10) BENSON, F. A., and CHALMERS, P. M.: 'Effects of Argon Content on the Characteristics of Neon-Argon Glow-Discharge Reference Tubes', *Proceedings I.E.E.*, Monograph No. 321 R, December, 1958 (**106** C, p. 82).
(11) WESTON, G. F.: 'Glow-Discharge Characteristics of Helium-Neon Mixtures', *British Journal of Applied Physics*, 1959, **10**, p. 523.
(12) FROUWS, S. M.: 'Paschen Curves in Neon-Argon Mixtures', Proceedings of the Third International Conference on Ionization Phenomena in Gases, Venice, 1957, p. 341.
(13) BENSON, F. A.: 'Initial Drifts in Running Voltage of Glow-Discharge Regulator Tubes', *Journal of Scientific Instruments*, 1950, **27**, p. 71.
(14) BENSON, F. A., and MAYO, G.: 'Effects of Ambient-Temperature Variations on Glow-Discharge Tube Characteristics', *ibid.*, 1954, **31**, p. 118.
(15) LOEB, L. B.: 'Mechanism of Cataphoretic Segregation in Inert Gas Glow-Discharges', *Journal of Applied Physics*, 1958, **29**, p. 1369.
(16) MEEK, J. M., and CRAGGS, J. D.: 'Electrical Breakdown of Gases' (Oxford, 1953).
(17) MASSEY, H. S. W., and BURHOP, E. H. S.: 'Electronic and Ionic Impact Phenomena' (Oxford, 1952).
(18) PENNING, F. M.: 'The Starting Potential of the Glow Discharge in Neon-Argon Mixtures Between Large Parallel Plates', *Physica*, 1934, **1**, p. 1028.
(19) JURRIAANSE, T.: 'The Influence of Gas Density and Temperature on the Normal Cathode Fall of a Gas Discharge in Rare Gases', *Philips Research Reports*, 1949, **1**, p. 407.
(20) SMITH, J.: 'Temperature Coefficient of Maintenance Potential of Glow-Discharge Voltage Stabilizer and Regulator Tubes', *British Journal of Applied Physics*, 1958, **9**, p. 122, and private discussion.
(21) COBINE, D. J.: 'Gaseous Conductors' (Dover, New York, 1958).
(22) VAN GEEL, C.: 'Untersuchungen von Gasentladungen mit Rüchsicht auf ihre dynamischen Eingenshaften und ihre Stabilität', *Physica*, 1939, **6**, p. 806.
(23) VERHAGEN, C. J. D. M.: 'Impedanzmessungen an Gasentladungsröhren', *ibid.*, 1941, **8**, p. 361.
(24) VAN GEEL, C.: 'Zelfinductie en Nawerking in Gasentladungen', Thesis submitted to the University of Delft, 1955.
(25) BENSON, F. A., and CHALMERS, P. M.: 'Impedance/Frequency Characteristics of Glow-Discharge Reference Tubes', *Proceedings I.E.E.*, Paper No. 3175 E, March, 1960 (**107** B, p. 199).
(26) BENSON, F. A., and CHALMERS, P. M.: 'Effects of Argon Content on the Characteristics of Glow-Discharge Tubes', *Electronic Engineering*, 1960, **32**, p. 218.

DISCUSSION ON THE ABOVE PAPER AND ON 'THE CORONA DISCHARGE AND ITS APPLICATION TO VOLTAGE STABILIZATION'* AND 'IMPEDANCE/FREQUENCY CHARACTERISTICS OF GLOW-DISCHARGE REFERENCE TUBES',† BEFORE THE ELECTRONICS AND COMMUNICATIONS SECTION, 8TH MAY, 1961

Mr. G. F. Weston: The investigation of commercial tubes at Sheffield has made available to the user information which could not be obtained from the manufacturer and has also shown the manufacturer where improvements could be made. Recently, Dr. Benson has extended his work to examining gas mixtures and cathode materials in tubes which he has constructed and pumped himself. I am carrying out similar work and should like to point out that each commercial tube is designed for a particular application. The gas pressure, composition and electrode geometry are governed by the current rating, potential requirements, etc. If a different gas is used, electrode geometry, gas pressure and processing schedule may well be chosen differently. There is a danger, therefore, in taking one commercially designed structure and drawing conclusions as to the best gas mixture to use, merely by changing the gas content. The paper by Dr. Benson and Mr. Burdett shows this very clearly. As example, for a uniform field the minimum breakdown potential depends on pressure times electrode spacing, *pd*. From Penning's work on neon-argon, however, we know that the minimum occurs at different *pd* values for different mixtures. Therefore the mixture giving the lowest breakdown will depend on the *pd* chosen. Also, from the paper by Mr. Cohen and Dr. Jenkins we see that, at breakdown, the field is by no means uniform for cylindrical geometries and therefore the Penning effect will not be the same as for parallel plates. Similarly, in the case of maintaining potential, it is seen from Table 1 that in some tubes an anode fall occurs (distinguished by the positive temperature coefficient). Were this not so, it is likely that the minimum would occur with a different percentage admixture. It should be stressed, therefore, that the results apply specifically to the geometry and pressure used, and that they may well be inapplicable to any other geometry or gas pressure. This is true in my own experiments, and I suggest that results with uniform fields over a wide *pd* range are more meaningful.

There is a statement on cataphoresis which does not seem to me correct. Cataphoresis does not take place until a discharge

* COHEN, E., and JENKINS, R. O.: Paper No. 3174 E, May, 1960 (see **107** B, p. 285).
† BENSON, F. A., and CHALMERS, P. M.: Paper No. 3175 E, March, 1960 (see **107** B, p. 199).

passes. How, then, does it affect breakdown potential? Again, if it affects the maintaining potential and not breakdown, would not the admixture for minimum maintaining potential be lower than for breakdown potential?

Our experience is that the spread between tubes depends on the processing schedule and if the optimum is used for each mixture it is independent of the gas used. We can, in fact, make close-tolerance tubes with pure gases if we take enough care. Usually the large spreads obtained with pure gases are because the gases are, in fact, not pure. This may be the reason for the spreads found by Dr. Benson. Looking at Table 1 one sees that the spread is not necessarily a minimum at 1% admixture, as suggested, but occurs at lower admixtures except in the case of neon plus xenon.

Considering the paper by Dr. Benson and Mr. Chalmers, the impedance of glow discharges at high frequencies has not previously been measured, and it is interesting to note that the results seem to fit the equations suggested by van Geel. Strictly speaking, however, van Geel's theory applies only at low frequencies; it assumes that when the voltage is changed, resulting in a change in current, the condition of the discharge is a function of these new values of voltage and current. The time it takes to get there is not taken into account; it is assumed that the frequency is low compared with this time. This would not be true at frequencies of the order of 1 Mc/s. The τ in van Geel's theory is the time-constant associated with the change in space charge. One would expect it, therefore, to have a value of the order of the build-up time of a discharge, i.e. 10^{-5}–10^{-4} sec. The delayed effects he considered would therefore have a time constant, T, greater than this. I suggest, therefore, that the τ for neon (Fig. 14) would be associated with 13 μs and not 0·02 μs as implied. Is van Geel's equation still valid for a delay effect where T is smaller than τ, and can the simplified equation (28) still be deduced for T much less than τ?

Mr. G. M. Ward: Can Dr. Benson throw some light on requirements I have for a tube to meet a certain performance? It is necessary to concentrate more on small physical size and wide swing of burning current. One could allow various constraints in a design such as this, but the importance of gas thermal conductivity, envelope temperature, etc., pose some interesting considerations.

Mr. J. Smith: I was particularly interested in Section 4.3 of the paper by Dr. Benson and Mr. Burdett, dealing with temperature coefficients. It has previously been suggested that positive coefficients could be due to impurities liberated by the heating of the envelope, but this explanation could not account for the fact that the maintaining voltage returns within a few minutes to its original value when the envelope is returned to ambient temperature. The authors, using tubes at fixed current and pressure, found that as the percentage of admixture was increased an anode fall developed and the temperature coefficient changed from negative to positive. They concluded that the formation of the anode fall caused the coefficient to become positive.

I have recently examined the same problem more directly using an adjustable-gap tube so that discharges with and without an anode fall could be examined in the same tube. The temperature coefficients for gap distances such that an anode fall was absent, partially developed and fully developed, were −8, +41 and +4 MV per deg C, respectively, confirming the authors' conclusion that the coefficient can change from a small negative value to a large positive value during anode-fall formation. In general, as the pressure is increased the maintaining voltage continues to increase after the anode fall is completely developed, but at a slower rate than during the anode-fall formation. It does not 'continue its original slope', i.e. decrease as stated by the authors. The low positive value of the coefficient given above for the fully developed case is therefore to be expected.

Concerning the paper by Dr. Benson and Mr. Chalmers, can transit times obtained from impedance measurements lead to estimates of the variation with pressure of the cathode-dark-space distance, and therefore of the product of pressure and this distance? It is well known that this product remains approximately constant at pressures sufficiently low for the cathode-dark-space distance to be measured visually or by electrical methods, but confirmation that this extends to pressures above, say, 10 mm Hg has not been obtained. This constancy is, however, implied in some theories of the cathode-fall region.

Mr. R. Lake: With regard, first, to the paper by Mr. Cohen and Dr. Jenkins, I have two questions. One parameter I believe to be of importance is repeatability of gap voltage and, to a certain extent, striking voltage, with repeated series of switchings. Will the authors comment? In Fig. 4, showing four tubes, the second one from the right has two holes in the cathode cylinder. What is their purpose?

In the paper by Dr. Benson and Mr. Burdett, the curves of running voltage against current sometimes show discontinuities which presumably correspond to the point where the cathode glow jumped outside the cathode cylinder, and are not to be confused with voltage jumps. Will the authors say whether they could measure any voltage jumps, or give any indication of the extent of jumps in various mixtures?

With regard to the paper by Dr. Benson and Mr. Chalmers, the measurements were made on all tubes in normal lighting conditions. Would measuring the tubes in total darkness have any effect on the impedance characteristics? Alternatively, would the addition of some radioactive source have any effect?

Regarding the temperature-coefficient discussion, I believe an anode fall of potential is always present in a corona tube, and at the same time these tubes always have a positive temperature coefficient.

Mr. R. P. Rowe: In using corona discharge tubes of the SC1/1600 type as a voltage reference for a series stabilizer in a regulated power supply, considerable trouble has been experienced because of their short life, approximately 200 h. Some failures seem to occur with recurrent switching.

The failures amount to something like 40 out of 60 valves. Could Mr. Cohen and Dr. Jenkins say if this is the normal expected life with recurrent switching?

Mr. W. E. Willshaw: Will Dr. Benson and Mr. Burdett give some indication of physical effects arising from the presence of admixtures which lead to the results described in their paper?

Mr. D. Rees: Will Mr. Cohen and Dr. Jenkins say what effect the proximity of the glass wall has on their discharge tubes?

THE AUTHORS' REPLIES TO THE ABOVE DISCUSSION

Dr. F. A. Benson and **Messrs. G. P. Burdett** and **P. M. Chalmers** (*in reply*): Mr. Weston quite rightly points out the dangers of assuming that the effect of varying a parameter, in this case gas content, measured under one set of conditions is the same when applied to another type of tube where conditions may be completely different. Since there are a large number of parameters which may be varied it was thought advisable to alter them in turn. The cylindrical structure was chosen so that results could be compared with much of our previous work; it was used originally because it is commonly employed com-

mercially, it is a simple construction to produce and the sputtering is fairly easily carried out, although this structure is not the best in this respect because a good deal of sputtered material is deposited on the anode or re-deposited on the cathode. Geometries producing uniform fields are not often, if at all, used in commercial tubes, probably because of electrode-cleaning and sputtering difficulties, and several attempts we made at producing parallel-plate tubes were not too successful on this account.

We have recently studied the characteristics of special tubes having molybdenum electrodes and filled with a 99·7% neon–0·3% argon mixture with five different cathode geometries. Gas pressures of 30, 40 and 50 mm Hg were used, and tubes with various cathode sizes and surface finishes, anode-cathode spacings and anode diameters were examined. The effects of cathode-geometry variations seem to be relatively unimportant. Initial drift, voltage/temperature, noise-voltage/current and impedance/frequency characteristics are not directly influenced by the electrode geometry. The shape of the running-voltage/current curve is somewhat dependent on geometry, but this is thought to be due to a combination of the low current density associated with the particular neon–argon mixture used and the influence of the cathode edges. More recent measurements on similar tubes with fillings of neon and other neon–argon mixtures confirm these results. If minimum running voltage is plotted against argon content for the cylindrical structures the curve agrees fairly well with that obtained by Jurriaanse, Penning and Moubis (Reference 8 in Benson and Burdett paper). There is a slight shift between the curves, possibly because these authors used plane electrodes.

The cataphoresis effect, as Mr. Weston points out, will be absent in the striking-voltage measurements, but it may have a slight effect on the minimum running voltage. More results would be necessary to determine whether the minimum striking and running voltages occur at exactly the same admixture percentage. Probably with the currents and pressures used the cataphoresis phenomenon might take a long time to show itself and be unimportant.

Although the tubes used for the present and previous investigations were manufactured under closely controlled conditions, spreads in results for tubes of a given type are quite large. Cathode-surface conditions and impurity variations from tube to tube due to unequal cleaning on sputtering are probably much more important here than the purity of the filling gas. Rapid clean-up of controlled added impurities in the gas filling which we have recently observed supports this. The surface structure of the cathode is probably largely determined by the sputtering and processing schedules in conjunction with the grain size and orientation of the metal used for its manufacture. There may be impurities in the grain boundaries of polycrystalline cathodes, or surface layers of impurities, particularly oxides, which cause spreads in results. Ahsmann and van Gelder* have recently studied the characteristics of tubes using single crystals of germanium, silicon and copper, and Benson and Rigg† have produced tubes with monocrystalline molybdenum cathodes. It has been found that running voltages of tubes with single-crystal cathodes and normal 'spectrographically-pure' gas fillings are much more reproducible than those of polycrystalline ones of the same material and are very constant with time.

It appears from Table 1 that minimum spread cannot be correlated with the percentage of admixture, but many other measurements, particularly on neon–argon tubes, have confirmed that the minimum spread coincides approximately with minimum running voltage. This may be due to the difficulty of measuring and controlling the amount of added admixture in the cases of the very low contents or, for large admixture contents, to the inability to age and sputter the tubes adequately. The high current densities, in conjunction with the electrode configuration, may have caused incomplete or uneven cleaning of the cathode surface and prevented total coverage of the glass walls with a molybdenum layer.

We do not agree with Mr. Weston's statement that van Geel's theory applies only at low frequencies. The only assumption made is that the amplitude of the superimposed a.c. signal is small. At high frequencies one need only consider the electron space-charge; as the voltage is changing there will then be a change in tube current which takes place in the order of an electron transit time. There will therefore probably be agreement between the form of the impedance loci obtained experimentally and the theoretical loci, provided that the frequency is small compared with the reciprocal of the electron transit time. Van Geel (Reference 6 in Benson and Chalmers paper, p. 93) discusses theoretical loci for both $\tau < T$ and $\tau > T$, and his equation is still valid when $T < \tau$. In order to fit the time-constants τ_1, τ_2, etc., to delay times in the discharge tube further measurements in the range 10–20 Mc/s are necessary. Eqn. (28) is an exact expression, not a simplified one, and applies to both the cases $T < \tau$ and $T > \tau$.

It is impossible to give a complete answer to Mr. Ward here, but we draw his attention to the advantages of using ceramic–metal tubes, hollow cathodes or even higher pressures. Lafferty* has described several types of small ceramic–metal tubes. In one, a stainless-steel wire coated with barium aluminide is press-fitted into a titanium disc forming the tube base. A ceramic washer, undercut to form a long internal leakage path, separates the anode from the titanium envelope, which serves as the cathode. On heating the structure to 1070°C the seals are made, the barium aluminide dissociates evaporating barium uniformly on to the cathode surface, and the anode is brazed to the titanium base by alloying between the titanium and the stainless steel. Lafferty points out that combinations of titanium (cathode material) and xenon (gas-filling) give a cathode fall of about 160 V, and he has used these materials to produce a high-voltage tube where several cells operate in cascade. A parallel-plane electrode geometry is used and 2000 V per inch of tube length has been obtained.

Culp and Koskos† have constructed a voltage reference tube with molybdenum electrodes in which a ceramic button insulates the anode and cathode, supports both and completes the envelope.

With a hollow cathode, for a fixed cathode area running voltage is independent of current because it is no longer a function of current density (see Fig. 51 in Reference 5 of Benson and Burdett paper; also the running-voltage/current curves in Reference 12 of Benson and Chalmers paper). As far as we are aware, no measurements of temperature coefficients or initial drifts have, however, been made on tubes with hollow cathodes.

The shape of the running-voltage/pressure curve described is only an assumption. We were mainly concerned with the formation of an anode fall, and Mr. Smith is probably correct in saying that after the anode fall is complete the curve continues to rise slowly and does not 'continue its original slope' as we assumed. In fact, Mr. Smith's comment is in agreement with Fig. 50 of Reference 5 in the Benson and Burdett paper.

* Ahsmann, G. J. M., and van Gelder, Z.: 'The Normal Cathode Fall on Single Crystal Cathodes', Proceedings of the Fourth International Conference on Ionization Phenomena in Gases (Uppsala, 17th–21st August, 1959), p. ID266. Also 'La chute cathodique normale pour des cathodes monocristallines', *Le Vide*, 1960, No. 87, p. 226.

† Benson, F. A., and Rigg, B.: 'The Stability of Reference Tubes using Monocrystalline Cathodes', *Electronic Engineering*, 1961, 33, p. 524.

* Lafferty, J. M.: 'A Process for making Clean Gas-Discharge Tubes', *Transactions of the Institute of Radio Engineers*, 1958, ED-5, p. 143.

† Culp, J. W., and Koskos, P.: 'A Ceramic-Metal Voltage Reference Tube', *ibid.*, 1957, ED-4, p. 144.

The frequency at which the tube reactance is zero is not necessarily correlated with the ion transit time, so transit times cannot be obtained from our experimental curves. The validity of the similarity rules can be investigated, however, by means of eqn. (30).

Voltage jumps of the type mentioned by Mr. Lake are frequently but not always encountered at low currents. When they are present they normally have a value of a few tenths of a volt. With single-crystal cathodes there are no such jumps, and so with polycrystalline material they may be due to the measured running voltage being an average of the running voltages of the different crystal faces exposed to the glow. If the glow moves to another part of the cathode then the crystal faces covered by the glow, and consequently the running voltage, will change. Tubes with cerium cathodes which we examined had running-voltage/current curves which exhibited voltage steps which increased in magnitude with the argon content in the main gas filling. These steps were probably largely due to the non-uniformities of the cerium-cathode surfaces employed.

All the impedance measurements were made with tubes in normal daylight. Measuring tubes in total darkness would not be expected to have any effect. Irradiation with light from a neon positive column can alter the number of metastable atoms in a neon–argon discharge tube and this causes a small change in running voltage which may affect the impedance at low frequencies. Reference tubes are, however, shielded by the sputtered layer so that the effect will be very small. We have made no measurements on tubes containing a radioactive source but we should not expect this to have any great effect on impedance characteristics.

With regard to Mr. Willshaw's question, the results have been discussed as fully as possible in the light of the physical effects present, e.g. Penning effect, density changes in cathode region, etc.

Mr. E. Cohen and **Dr. R. O. Jenkins** (*in reply*): In reply to Mr. Lake's questions, the corona discharge gives good repeatability of gap and striking voltage on repeated switchings. There is a relatively small difference between the striking and maintaining voltages, and the glow always covers the anode completely so that the variations in glow position often seen in the glow discharge tube are absent. The holes in the cathode cylinder are there to assist processing of the tube.

In reply to Mr. Rowe, the premature failures he experienced were unlikely to be caused by recurrent switching. The fault was probably a weak pin seal from which some early tubes suffered. On insertion into a stiff holder, failure of the seal allowed air to leak in slowly, causing the short life. Switching has little effect on the life, which should normally be some thousands of hours.

In reply to Mr. Rees, the proximity of the glass wall has little effect on the operation of the corona discharge tube. In practice, the discharge is completely contained within the cathode cylinder and end insulators. The only phenomenon which might possibly be associated with the glass wall is that the minimum current for steady operation when tubes are in series does appear to be altered by the presence of conductors at different potentials near the tube.

621.397.621 : 621.385.832

The Institution of Electrical Engineers
Paper No. 3468 E
Feb. 1961

PROGRESS REPORT ON THE DEVELOPMENT OF A PHOTO-ELECTRIC BEAM-INDEX COLOUR-TELEVISION TUBE AND SYSTEM

By R. GRAHAM, M.Eng., J. W. H. JUSTICE, and J. K. OXENHAM, M.A.

(*The paper was first received* 20*th May, and in revised form* 21*st October*, 1960. *It was published in February*, 1961, *and was read before the* ELECTRONICS AND COMMUNICATIONS SECTION 1*st May*, 1961.)

SUMMARY

The paper is a progress report on the development of an index colour-television display using photo-electric indexing, and outlines some of the basic problems associated with index systems. The effect of cross-modulation between the writing and index signals is examined and a method of overcoming this cross-modulation is suggested. The effect of the delay around the index loop and its relationship with the horizontal time-base linearity is discussed, and a preferred circuit arrangement which relaxes the tolerances on this linearity is described. Some comparisons are drawn between index and other colour-television displays, and details are given of the practical verification of the system described. The present system may lead to a simplification of the colour display tube.

LIST OF PRINCIPAL SYMBOLS

V_s = Complete writing signal.
V_m = Monochrome component of writing signal.
V_y = N.T.S.C. luminance signal.
V = Modified writing signal.
V'_R, V'_G, V'_B = Gamma-corrected colour signals.
V_c = Amplitude of chroma writing signal.
ω_c = Angular frequency of transmitted chroma signal.
ω_{cR} = Angular frequency of transmitted reference signal.
ω_s = Angular frequency of writing signal.
ϕ = Phase of chrominance signal.
θ_{max} = Maximum phase error of the index signal.
γ = Index of power law of the tube transfer characteristic.
p = Pitch of the complete colour sequences.
d = Diameter of focused spot measured at the half-brightness level.
$F(t)$ = Phase perturbation of the index signal.
$\Delta\omega_s$ = Change in the angular writing frequency.
I_a = Cathode-ray-tube current.
I = Complete index signal.
A_r = Fourier coefficients in I_a.
B_r = Fourier coefficients in I.
T = Phase/frequency slope.
$\phi_a - \phi_f$ = Phases of index and writing signals.

(1) INTRODUCTION

At the inception of the present work[1, 2] in 1954, there were two known practical displays of the single-tube variety, the multi-gun shadow-mask tube[3, 4] and the single-gun focus-mask or beam-switching tube.[5] Both these types of display tube are complex because the method of phosphor selection involves a mechanical structure inside the glass envelope of the tube, which must be accurately made, positioned, and aligned electron-optically with the phosphor pattern of the screen.

With the multi-gun tubes there are two main problems. First, there is the problem of matching the transfer characteristics of the three guns, and secondly, the problem of superimposing the three colour images over the whole screen area. Both these problems can be avoided by using a single electron gun and sequentially exciting the three phosphors.

A well-known example of a display in which the phosphor patterns are sequentially excited is the beam-switching tube, in which voltages are applied to a grid of wires placed near to the screen in order to switch the electron beam from one colour pattern to another, but a much simpler way of sequentially exciting the three phosphor patterns is to use the normal television scanning to perform the switching. For example, if, in a normal monochrome television tube, the screen were replaced by a screen consisting of three sets of equi-spaced vertical lines, one red, one blue and one green, the electron beam would excite the three colours sequentially owing to the action of the horizontal scan.

A feature of displays of this sort is the ease with which it is possible to produce a monochrome picture. If a monochrome signal is applied to this type of tube, the three phosphors are equally excited since the phosphor pattern is finer than the finest pattern reproducible by the television system. In fact, the situation is similar to that which occurs in a normal monochrome television tube where the 'white' picture is obtained from a mixture of colour phosphors.

To produce a colour picture on any of these systems it is necessary to have some means of indicating the position of the beam in the red–green–blue stripe sequence, or colour triad. Many ideas have been suggested and feature in the patent literature from 1937 onwards.[6, 7] They consist mainly of the inclusion in the screen of some pattern related to the colour–phosphor sequence, this pattern causing an 'index' signal to be produced when it is scanned by an electron beam. Various physical properties have been suggested for producing this index signal, including secondary emission, photo-electric and conduction phenomena.

The early patents, although disclosing the index principle, do not discuss the difficulties in using the index or positional information. The first practical solution was demonstrated in 1956 to the C.C.I.R.[8–10] and involved the use of secondary-emission indexing[11–13] and two electron beams to separate out the index information from the colour-writing signals.

There are many possibilities in the choice of the number and arrangement of the sets of phosphor stripes and, in particular, whether they are placed horizontally or vertically.[14] Any system employing vertical phosphor stripes usually has a counterpart in a system using horizontal phosphor stripes, and the requirements and problems associated with these two classes of systems are similar in nature, e.g. spot size, time-base linearity, cross-modulation of the index signal by the writing signal, etc.

It is the purpose of the paper to discuss some of the problems associated with index tubes and systems, and to describe another approach to the problem of separating out the index signal and the subsequent processing to produce a suitable writing signal.

The authors are with Sylvania-Thorn Colour Television Laboratories, Ltd.

This new approach has been specifically directed towards the simplification of the colour-display tube.

(2) PICTURE WRITING

As the electron beam is scanned across the red, blue and green phosphor stripes, the signal applied to the grid of the tube must in turn represent the gamma-corrected red, blue and green components of the picture to be reproduced. This condition can be satisfied by applying to the grid of the tube a signal of the form

$$\left.\begin{aligned} V_s &= V_m + V_A \cos \omega_s t + V_B \sin \omega_s t \\ \text{where} \quad V_m &= \tfrac{1}{3}(V'_R + V'_G + V'_B) \\ V_A &= \tfrac{1}{3}(2V'_R - V'_G - V'_B) \\ V_B &= \frac{1}{\sqrt{3}}(V'_B - V'_G) \end{aligned}\right\} \quad . \; . \; (1)$$

assuming that the red stripes are scanned when $\omega_s t = 2\pi n$ and that the blue stripes will therefore be scanned when $\omega_s t = 2\pi n + 2\pi/3$ and the green stripes when $\omega_s t = 2\pi n + 4\pi/3$.

One can easily verify that the signal V_s is satisfactory by substituting the appropriate values of $\omega_s t$ into eqn. (1), which can be rewritten in the form

$$\left.\begin{aligned} V_s &= V_m + V_c \cos(\omega_s t - \phi) \\ \text{where} \quad V_c &= \sqrt{(V_A^2 + V_B^2)} \\ \text{and} \quad \phi &= \arctan \frac{V_B}{V_A} \end{aligned}\right\} \quad . \; . \; . \; (2)$$

In eqn. (2) the ratio V_c/V_m determines the saturation of the colour produced and ϕ determines the hue.

In eqns. (1) and (2) no account has been taken of the effects of the finite size of the phosphor stripes and the focused spot. As the width of the stripes and the size of the focused spot are increased the saturation of the colour produced is decreased without undue change in hue. However, this effect can be offset to a great extent by increasing the value of V_c/V_m.

A full account of the colour errors produced in a display of this type is given elsewhere.[15] In this Reference it is shown that reasonable colour accuracy can be obtained with a screen structure in which the width of the phosphor stripes is equal to the spacing between the stripes and with a spot diameter (measured to half brightness) of the order of one-third the pitch of the colour sequences.

As mentioned above, the hue of the reproduced colour is determined by ϕ. Experiments have shown that an error of 10° in ϕ produces a discernible change of hue. Since the number of colour sequences required in the width of the picture is of the order of 300, a 10° error in ϕ corresponds to a displacement error of approximately 10^{-4} picture widths or about 0·002 in on a 21 in tube. It seems unlikely that this order of accuracy of scanning could ever be achieved in a practical receiver without some other control. The difficulty of obtaining this degree of accuracy of scanning has already been discussed.[16] The system described in this Reference is another variation of the sequential-writing principle in which the phosphor stripes are horizontal and in which a small wobble signal is added to the frame scan. This transfers the scanning accuracy problem from the line scan to the frame scan, but even then no practical solution could be found which did not involve some other control signals.

(3) BEAM INDEXING

The principle of beam indexing is to include in the screen of the cathode-ray tube some means of generating an index signal which is a function of the instantaneous position of the beam. The index signal can be obtained from a set of stripes, parallel to the phosphor stripes, of a material which differs from the rest of the screen either in its secondary or photo emission. After amplifying the output of a suitable pick-up device, a sinusoidal index signal is obtained, the phase of which contains information about the relative position of the scanning beam. This index signal can then be used to control either the scanning or the writing signal to maintain the two in register.

If the scanning is to be controlled, the phase of the index signal is compared with that of the carrier from which the writing signal is derived and the error signal obtained is used to control the scan. As this servo control must be fast-acting to be effective, there is some practical difficulty in applying the error signal to the scanning. For this reason, the preferred method of operation is to derive the writing signal from the index signal and leave the scan uncontrolled. The system then becomes an open-loop control provided that the index signal is not cross-modulated by the writing signal.

(4) THE CROSS-MODULATION PROBLEM

Fig. 1 shows the index signal obtained from a simple index structure in which there is one index stripe coincident with each red phosphor stripe; (*a*), (*c*) and (*e*) show the signals obtained from parts of the picture which are white, red and cyan, respectively.

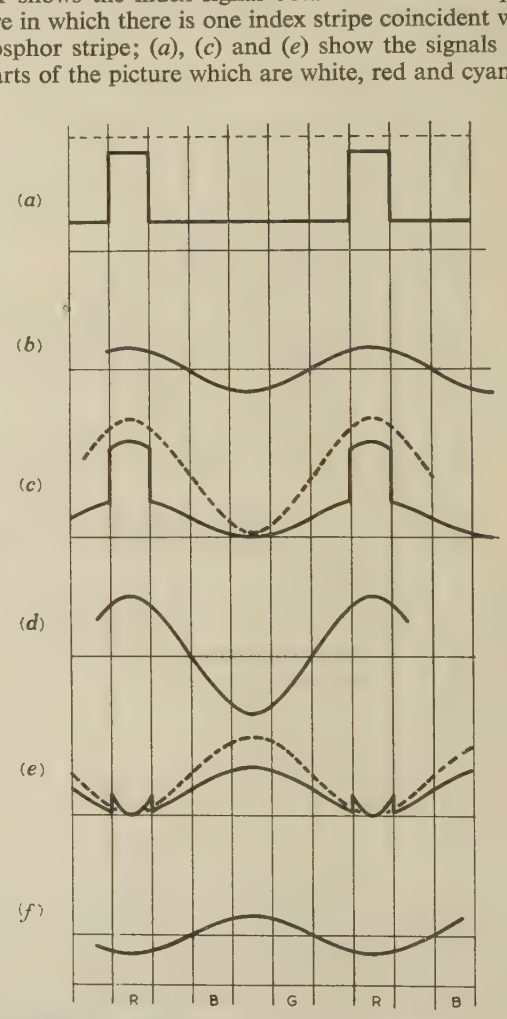


Fig. 1.—Index signal obtained from pictures of different colours.

--- Writing signal. ——— Index signal.

(*a*) Index signal from a white picture.
(*b*) Fundamental component of waveform (*a*).
(*c*) Index signal from a saturated red picture.
(*d*) Fundamental component of waveform (*c*).
(*e*) Index signal from a saturated cyan picture.
(*f*) Fundamental component of waveform (*e*).

tively, and (*b*), (*d*), and (*f*) the fundamental components of these signals after filtering. It can be seen that:

(i) The amplitude of the fundamental component of the index signal is a function of the writing signal. This effect is of little importance since a limiting amplifier can be used.

(ii) The phase of the fundamental component of the index signal can become more dependent on the writing signal than on the position of the beam when an attempt is made to produce colours of high saturation. Since the writing signal is itself derived from the index, it is possible to reach a condition of instability in which the system oscillates at a frequency near to the index frequency.

Such a simple system would obviously not work because of this cross-modulation. One method of overcoming this effect is to use a carrier, as in the 'Apple' system,[8] and to place the index information outside the frequency band occupied by the writing-beam currents. In this system, two electron beams are used, a writing beam and a low-current pilot beam which is modulated with a high-frequency carrier.[13] The index stripes are of a material having a coefficient of secondary emission different from the rest of the screen, and the secondary electrons are collected by a suitable electrode in the bulb. The current at the collector electrode consists of a high-frequency carrier amplitude modulated by the index signal, together with some lower-frequency components due to the writing-beam currents. One of the sidebands of the modulated carrier is selectively amplified and heterodyned with the original carrier wave to produce an index signal. This signal is not affected by the writing signal because the sideband frequency used lies outside the frequency spectrum of the writing-beam current.

Unfortunately, this system suffers from a number of practical difficulties due to the use of two beams and because the finite flight time of the secondary electrons requires the index pattern to be a different shape from the phosphor pattern, but the idea of placing the index signal in a part of the frequency spectrum where there are no writing-beam components does suggest an alternative approach to the problem of reducing cross-modulation in a single-beam tube.

(5) NON-INTEGRAL SYSTEMS

One way of producing an index signal outside the spectrum of the writing signal with a single-beam tube is to use an index structure having a periodicity other than that of the phosphor sequences.[17] Since the writing signal must be derived from the index signal, there must obviously be some frequency relationship between the two. In general, the index frequency could be m/n times the writing frequency. For various reasons it is desirable to put gaps between the phosphor stripes[18] and it is convenient to be able to place the index stripes in these gaps, leading to a value of $m = 3$. Practical considerations dictate that the writing-signal frequency is about twice the maximum video frequency, and so values of n above about 4 (for $m = 3$) would bring the index frequency very near or within the spectrum of the V_m component of the writing signal. Even $n = 4$ is marginal, since the frequency spectrum of the V_m signal is doubled by the approximately square-law characteristic of the tube.

In addition to this contamination of the index signal by the V_m signal and its harmonics, there is also some cross-modulation caused by the mth harmonic of the writing signal in the beam current and the $(n - 1)$th and $(n + 1)$th harmonics of the index waveform. This spurious signal is, of course, at the same frequency as the true index signal, but the phase is a function of the phase of the writing signal (see Section 13.1). It therefore causes a phase error in the index signal, and Fig. 2 shows how the maximum phase error, θ_{max}, of the index signal varies with the spot size for the three systems ($m/n = 3, \frac{3}{2}, \frac{3}{4}$). Some

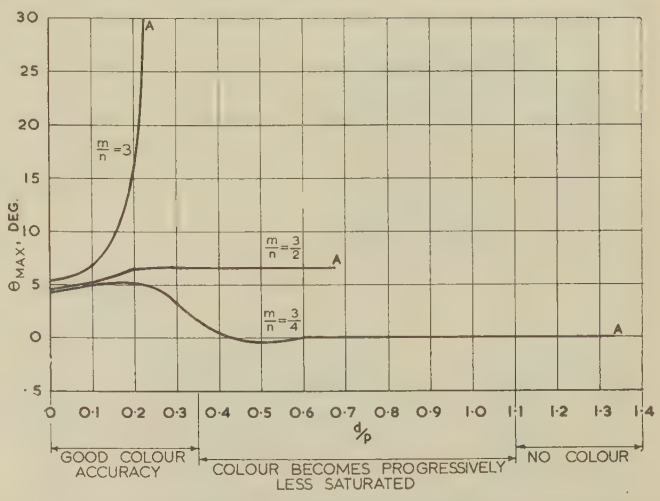


Fig. 2.—Maximum phase error of the index signal as a function of spot size for three non-integral index system $m/n = 3$, $m/n = \frac{3}{2}$ and $m/n = \frac{3}{4}$.

indication of the effect of spot size on the reproduction of colour is also given. The curves of Fig. 2 were obtained according to Section 13.1 and the following assumptions were made:

(*a*) That the tube has a square-law characteristic (i.e. $\gamma = 2$).

(*b*) That the width of the index stripes is the same as that of the phosphor stripes $= p/6$.

(*c*) That the distribution of electrons across the focused spot is such that the number of electrons in an elemental strip parallel to the phosphor stripes is proportional to the square of the cosine of the distance of the strip from the centre of the spot. This approximates quite well to measured distributions.

(*d*) That the drive applied to the tube is equivalent to that used to produce a saturated colour. This represents the worst case, since the cross-modulation always increases with saturation.

The points A on the curves are the points at which the index signal becomes unusable (see Section 13.1). Although the three systems have similar performances for very small spot sizes, there are marked differences as the spot size increases, and, in fact, when $m/n = 3$ the spot-size requirements for indexing are more stringent than those for good colour reproduction. For this reason, this system is not used. Although θ_{max} is less for $m/n = \frac{3}{4}$ than for $m/n = \frac{3}{2}$ this is not the only consideration, and, as mentioned above, there is a possibility of cross-modulation by the video signal when $m/n = \frac{3}{4}$. Experiments have shown that $m/n = \frac{3}{2}$ is probably the best compromise.

It is interesting to note that, for many of the systems that have been analysed, there are dual systems with horizontal stripes and vertical wobble (Fig. 3 shows an example of this duality), but in the case of the $\frac{3}{2}$ system, there appears to be no such simple dual system.

One difficulty with all these non-integral systems is that, in order to derive the writing signal from the index signal, it is necessary to perform a frequency division, and when this is done the resulting signal can have one of m phases, only one of which will produce the correct colour. To overcome this difficulty, 'run-in' index stripes are provided on the left-hand side of the tube, which produce an auxiliary index signal at a frequency which is a submultiple of the writing frequency. This auxiliary index signal is used to start the divider circuit in the correct phase at the beginning of each line scan. Once the divider has started, it is essential that the index signal is maintained until the end of the line, to prevent the divider stopping and restarting in the wrong phase. To ensure that this happens, the minimum black level of the picture is limited.

The authors have found that the secondary-emission type of

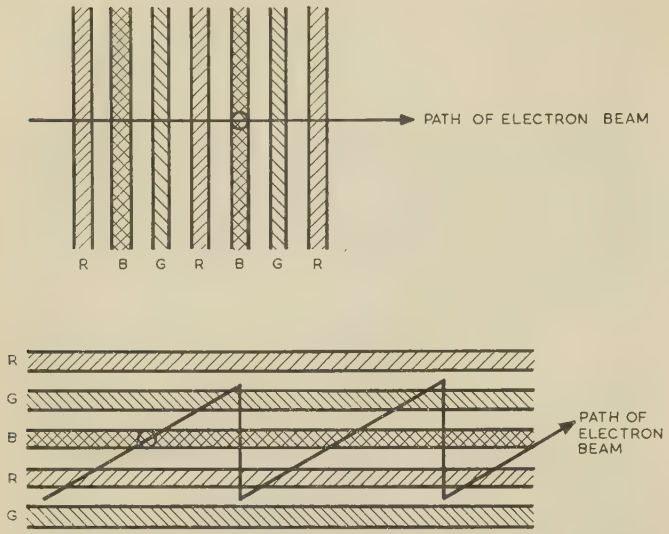


Fig. 3.—An example of the duality between systems employing vertical and horizontal phosphor stripes, and the method of obtaining sequential writing in the latter case.

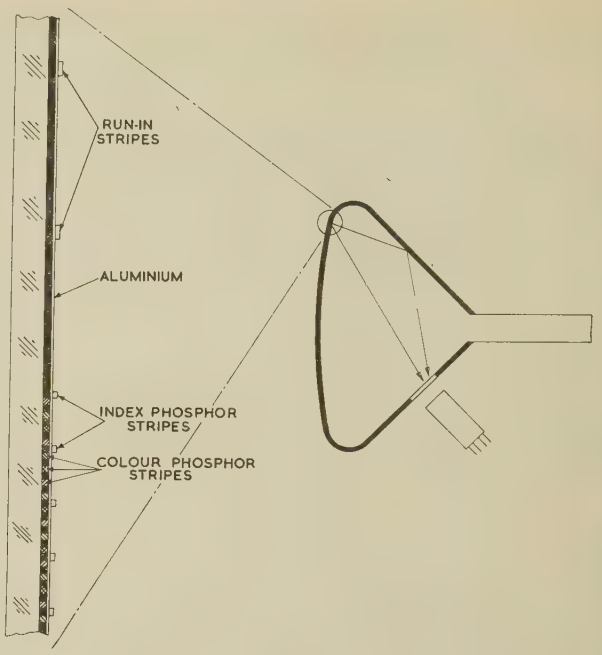


Fig. 4.—Horizontal sectional view of the tube, showing the positio of the photo-multiplier and a section of the face of the tube wit the index and colour phosphor stripes for a $3\omega_s/4$ system usi $\omega_s/4$ run-in stripes.

indexing used in the 'Apple' system does not work very satisfactorily with single-beam tubes since the relationship between the index yield and tube beam current is unpredictable. It is believed that this is due to space-charge effects at high beam currents. These effects are not noticeable on the 'Apple' tube, since the two electron beams strike the screen at points a sufficient distance apart from the space charge due to the writing beam to have a negligible effect on the secondary emission from the pilot beam.

Fortunately, due to the lower index frequency of the $\frac{3}{2}$ or $\frac{3}{4}$ systems, it is possible to obtain an index signal by photo-indexing, so eliminating the variation of flight time and ensuing colour errors of the secondary-emission system.

(6) PHOTO-INDEXING

The phase of the photo-index signal contains the positional information required for the operation of the system, and any noise in this signal will generally degrade the quality of the picture. The prime source of noise is the photo-cathode of the photo-multiplier, and this noise is proportional to the square root of the light collected by the photo-cathode. It is desirable, therefore, that the photo-indexing light is maximized whilst all other light is reduced to a minimum.

A logical position for the index stripes is on the interior face of the aluminium film, with the photo-multiplier placed behind the cone of the tube, as shown in Fig. 4. In this position the film acts as a filter preventing the light from the colour phosphors reaching the photo-cathode, and also as a reflector for both the index and the colour phosphor stripes. In addition, the light yield of the index stripes is enhanced by internal reflection in the tube, causing the aluminized cone and face-plate to act as a light-integrating 'sphere'. Experiments show that an improvement of between 2 : 1 and 5 : 1 can be achieved, depending upon the choice of window position for monitoring the signal and the location of the spot on the screen.

The photo-index phosphor is chosen to have a high ultra-violet light content so that the light signals from the index phosphors can be optically filtered from the colour luminescence and from other sources of stray light. It must also have a very short persistence, or 'modulability' at the index frequency, which is of the order of 10 Mc/s, and the index phosphor in current use is P16 (calcium-magnesium silicate activated b cerium).

This phosphor has a practical frequency limit of about 15 Mc/ above which the a.c. component of the light signal falls o rapidly. This will degrade the signal/noise ratio of the inde signal and is one of the reasons why photo-indexing with a index tube employing a separate pilot carrier beam is not a ver practical proposition.

Although placing the index phosphor stripes on the interio face of the aluminium has considerable advantages for th operation of the indexing system, it does pose considerab technological problems in the manufacture of the tube and place some restrictions on the method of application of the phospho stripes.

A number of other approaches to the screening problem a under investigation and will be described in a companio paper.

(7) CIRCUIT CONSIDERATIONS

The requirements of the tube from its associated circuits a that good focus in the horizontal direction, sufficient to resolve single phosphor stripe, is maintained all over the usable scre area, and that it is supplied with the correct drive waveform its modulating electrodes.

The problem of achieving a sufficiently good focus quali depends upon a large number of factors, such as the design of t electron gun, focusing and deflector coils, deflection angle, t particular compromise of brightness and phosphor-triad spaci employed in the tube, e.h.t. regulation, etc. Without going in details, it has been possible to meet this requirement in a 14 rectangular tube of 70° deflection angle with a triad pitch 0·031 in (32 triads/in) and a highlight brightness of 20 ft-lambe The e.h.t. regulation required was $\pm 1\%$, and focus modulati was applied in both the horizontal and vertical directions.

The form of writing signal applied to the modulating electro of the tube is again a compromise between the ideal and wh can be accomplished without undue circuit complexity. Th

shows a similarity with the beam-switching tube operated at a high switching rate. The colour and luminance errors produced by a simple writing signal of the form

$$V = \tfrac{1}{3}(V'_R + V'_G + V'_B) - k_1 V_c + V_c \cos(\omega_s t - \phi) \quad (3)$$

where

$$V_c \cos(\omega_s t - \phi) = K_2 \left[V'_R \cos \omega_s t + V'_G \cos\left(\omega_s t + \frac{2\pi}{3}\right) + V'_B \cos\left(\omega_s t + \frac{4\pi}{3}\right) \right] \quad (4)$$

have already been described,[15] and methods of obtaining a signal in this form are given below.

The tube also supplies information in the form of an index signal whose amplitude is a function of the mean beam current at any instant and the position of the beam, and whose phase is indicative of the instantaneous position of the spot in a colour triad. This last statement can be rephrased by saying that the index frequency is proportional to the rate at which the spot traverses the index stripes.

This index signal is processed and combined with the transmitted chroma instructions to provide the writing signal, V_s. The total variation in amplitude of the index signal can be as much as 500 : 1, and it is convenient to limit this index signal to a constant amplitude, still preserving the all-important phase information. This can be done in several ways, as follows:

(*a*) Conventional limiting techniques can be employed.
(*b*) The photo-multiplier can be operated in a non-linear mode (space charge).
(*c*) The video-drive signal to the tube can be applied to the amplifier to vary its gain to compensate for the variation in tube current. This gives an open-loop control and can easily be achieved by applying the video drive to one or more of the dynodes in the photo-multiplier.

In practice, either methods (*b*) or (*c*) can be used to lower the variation of index signal amplitude from 500 : 1 to about 20 : 1, followed by a limiting amplifier using conventional techniques.

The instantaneous frequency of the index signal is dependent upon a large number of factors, such as the accelerating potential of the electron gun (which will vary because of the finite impedance of the e.h.t. supply), and if cathode modulation is employed, the rate of change of current through the deflector coils (a function of the horizontal time-base linearity and width), the linearity and raster geometry of the deflector coils, the inside contour of the face-plate of the tube, damped oscillations in either the deflector coils or the horizontal output transformer, etc. It is therefore important to realize what effect any variations of index frequency will have on the accuracy of colour reproduction.

Section 13.2 gives a simple analysis of the effect of time delay in the index loop when the index frequency varies. Eqn. (17) and Fig. 10 show that the phase error caused by a constant change in the writing frequency, such as might be occasioned by a change in horizontal time-base width or e.h.t., is given by

$$\mathrm{F}(t) = t\Delta\omega_s$$

or

$$\phi_e = \mathrm{F}(t) - \mathrm{F}(t - T) = T\Delta\omega_s \quad . \quad . \quad . \quad (5)$$

Inserting some practical figures of $\omega_s/2\pi = 7$ Mc/s, $T = 1$ microsec, and $\Delta\omega_s/2\pi = 70$ kc/s representing, say, a change in width of 1%, then $\phi_e = 25{\cdot}2°$ and this phase error would give a quite discernible change in hue. When all the factors mentioned above which might affect the index frequency are considered, a $\pm\frac{1}{2}\%$ tolerance in index frequency is a difficult requirement to meet.

By rearranging the circuit (Fig. 10) as shown in Fig. 11, some advantage can be gained. In the phase domain, the final mixing process subtracts the phases of the $3\omega_s/2$ and $\omega_s/2$ signals. Thus, if it can be arranged by suitable choice of coupling circuits that the phases of the $3\omega_s/2$ and $\omega_s/2$ signals vary by the same amount and in the same direction, the net phase error will be zero.

This is somewhat of an over-simplification and takes account only of the static and not the dynamic case. However, by inserting $\mathrm{F}(t) = t\Delta\omega_s$ in eqn. (25), it is obvious that there is no phase error for a fixed index-frequency change.

This fact is evident by comparing eqns. (18) and (26), eqn. (26) being independent of $\mathrm{F}'(t)$. When the dynamic case is considered, this circuit arrangement has, however, certain disadvantages for rapidly changing variations in index frequency, as might occur if there are any damped oscillations in the deflector-coil current. This can be verified by putting $\mathrm{F}(t) = \cos pt$. For certain values of p and T, the circuit of Fig. 11 can give worse phase errors than that of Fig. 10.

There are two methods which can be used to produce a writing signal of the form given by eqn. (1) when dealing with a N.T.S.C.-type signal.

The first is to demodulate the transmitted chroma signal in two directions to give two colour-difference signals, and then to use these to remodulate the writing-signal carrier. This method is shown schematically in Fig. 5, in which the transmitted chroma

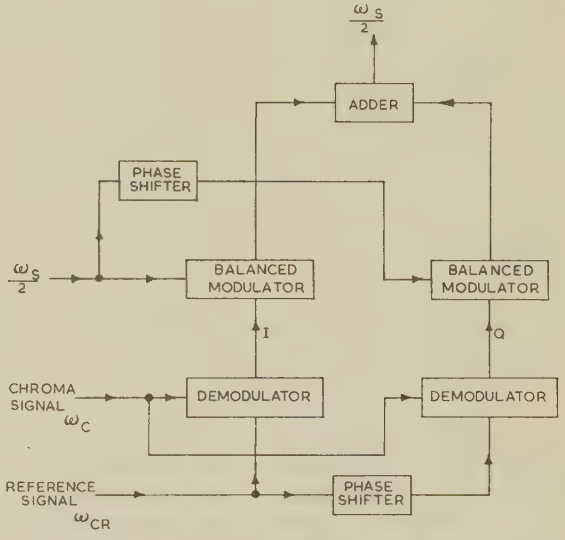


Fig. 5.—Demodulation-remodulation method of obtaining the colour-writing signal.

signal is demodulated by two quadrature reference signals and remodulated in balanced modulators. It is convenient to make one of the demodulated signals the $V_y - V_m$ signal, since this can then be used to convert the V_y signal to the V_m signal.

The other method is to use a heterodyne technique. The transmitted N.T.S.C. chroma signal is first converted into an equi-angular signal by an elliptical amplifier,[19] and then the chroma information is transferred to the writing-signal carrier by double mixing, as shown in Fig. 6. In this circuit, the chroma signal ω_c is first mixed in mixer 1 with the index signal $3\omega_s/4$ and the heterodyne signal at a frequency $3\omega_s/4 - \omega_c$ is selected. This signal will have the same chroma amplitude and phase information as the original ω_c chroma signal. The reference signal ω_{cR} is mixed in mixer 2 with the divided-index signal $\omega_s/4$ and the heterodyne signal at a frequency of $\omega_s/4 + \omega_{cR}$ selected. In a similar manner, this signal will incorporate

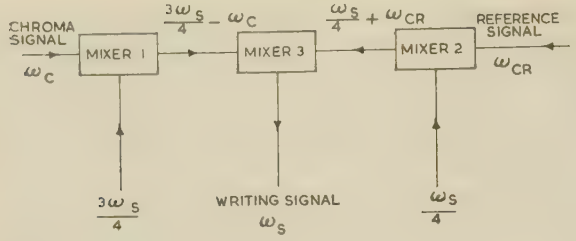


Fig. 6.—Heterodyne method of obtaining the colour-writing signal.

the phase information of the reference signal. Finally, the outputs of mixers 1 and 2 are combined in mixer 3, where the heterodyne signal at a frequency of ω_s is selected. This signal will be amplitude and phase modulated in the same way as the original chroma signal ω_c and so will be of the form given by eqn. (1). There are many variants of this technique, using $\omega_s/4$ and $3\omega_s/4$, or $3\omega_s/2$ and $\omega_s/2$ carriers, etc.

This last scheme has the disadvantage that the mixing action produces many unwanted frequencies which can produce annoying beat patterns on the face of the tube, and the choice of the carrier frequencies is somewhat governed by the ease with which some of these unwanted frequencies can be filtered out, yet still preserving the required bandwidth in the chroma- and writing-frequency circuits.

(8) SPECIAL CIRCUITS

Most of the circuits used in the work on system engineering have been of a conventional nature and use techniques common in current colour-television practice, but certain special circuits have been developed to meet the needs of the index tube. These are described below with reference to some of the authors' early work, in which an index frequency of $3\omega_s/4$ was chosen in preference to $3\omega_s/2$ because of the smaller cross-modulation due to colour-writing-signal harmonics, less stringent spot-size requirement for indexing, and because of the improved light yield from the index phosphor at lower frequencies. It does, of course, suffer from greater cross-modulation from the V_m signal, as already mentioned in Section 5.

(8.1) Horizontal Time-Base

To achieve the very high accuracy and linearity required, a conventional efficiency circuit was used, but with the output valve operated in a class A rather than the usual class C condition. The efficiency diode thus conducted throughout the scan. To compensate for the resistive losses in the circuit, an adjustable correction voltage was introduced in series with the damper diode, the voltage being approximately a sawtooth in shape. This sawtooth voltage is produced by a diode and capacitor combination from a voltage pulse obtained from the horizontal output transformer, as shown in Fig. 7. By this means, a linearity, referred to the phosphor-stripe spacing, of 0·5% was achieved. Well-known techniques were used to balance out transformer and deflector-coil oscillations.

Two forms of picture-width control were employed. The index frequency was measured in a discriminator circuit, the d.c. output of which was fed back to the grid of the time-base output valve in such a way as to maintain the average index frequency constant. This system therefore compensated for variations in mains voltage, warm-up drifts, etc. The deflector coils were chosen to give the smallest possible aberrations, and, for expediency, coils suitable for a shadow-mask tube were used. As a result, the raster appearing on the tube face was slightly pincushion in shape. To keep the index frequency constant over

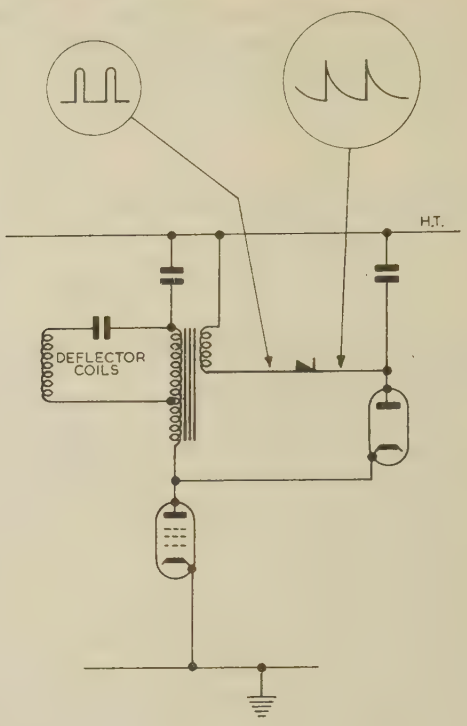


Fig. 7.—Schematic of the method used of compensating for the resistive losses in the line-scan circuit.

the frame scan, the picture width was modulated by a suitably phased parabolic waveform at frame frequency.

(8.2) The Divide-by-Three Circuit

In this particular system, the run-in stripes were chosen to be at an angular frequency of $\omega_s/4$ and the index stripes at $3\omega_s/4$ (Fig. 4). The width of the run-in stripes was made larger than that of the index stripes to enhance both the $\omega_s/4$ and $3\omega_s/4$ signals from them. Fig. 8 shows a block diagram of the circuit

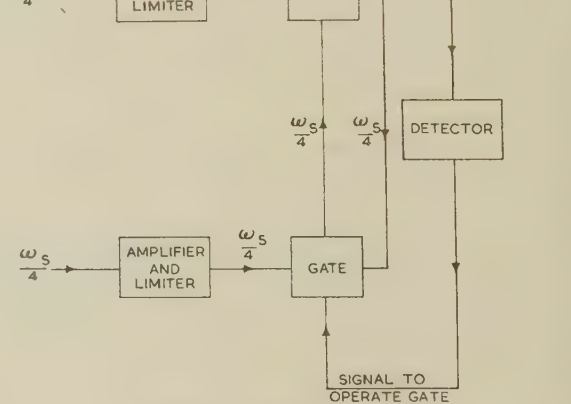


Fig. 8.—Divider circuit.

At the start of the horizontal scan during the run-in period the index structure provides both $\omega_s/4$ and $3\omega_s/4$ signals, which are amplified, limited and applied to the divider circuit. The divider stage acts as a gate during this run-in period, conducting only during the positive half-cycles of the two sine waves. It gives out pulses at an angular frequency of $\omega_s/4$, which are then filtered and fed back to the $\omega_s/4$ input of the divider. If the

phase shift and gain are suitably adjusted, the circuit will continue to run on the $3\omega_s/4$ input only, giving an $\omega_s/4$ output as long as the $3\omega_s/4$ input signal continues, i.e. until the end of the scan. The bias conditions on the divider are such as to prevent it starting unless both $\omega_s/4$ and $3\omega_s/4$ signals are present together, thus ensuring against false starting in the wrong phase. The frequency and phase of the output of the divider during the scan period will then be determined by the uncontaminated $3\omega_s/4$ signal only.

When the divider has started, its output is rectified and applied as bias to the $\omega_s/4$ amplifier to prevent any further signals from reaching the divider. This is of considerable importance since writing signals are being applied to the grid of the cathode-ray tube, which, together with the $3\omega_s/4$ index structure, would produce an amplitude- and phase-modulated signal. This signal could produce phase errors in the divided output.

The divider continues to divide throughout the scan, and ceases to divide when the $3\omega_s/4$ signal ceases during the flyback. During this time, the prohibitive bias is removed from the amplifier, which will then pass signals at the start of the next horizontal scan.

It is important that the $3\omega_s/4$ index signal persists throughout the scan, otherwise the divider may cease to operate and will not restart until the start of the next scan. For this reason, the tube beam current is always maintained at some preset low value, even during the 'black' portions of the picture.

Finally, a block diagram of the complete system is shown in Fig. 9, indicating how the various component parts of the circuit are related.

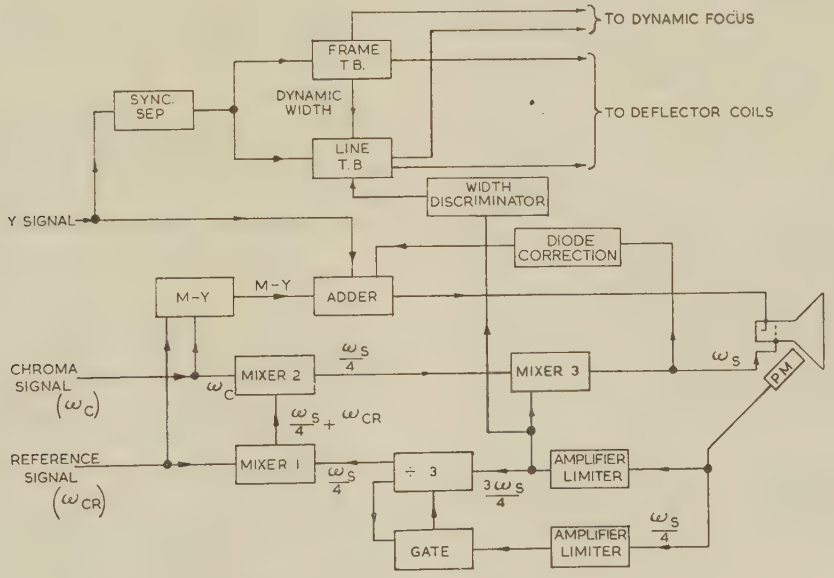


Fig. 9.—Complete index system.

(9) PHOTO-MULTIPLIER

The photo-multipliers used so far in experimental work have been of the conventional head-on type, which are readily available.

The electrical requirements of the photo-multiplier are not very exacting. The photo-cathode current for a photo-cathode area of 1 in^2 is about 0·02 μA for a tube beam current of 10 μA, so that a relatively modest gain of about 5000 is required. Furthermore, linearity of anode current with light input is not necessary or even desirable (see Section 7).

The authors have done no work on the economics of this device. Since one photo-multiplier is used for each receiver, it must be put on a mass-production basis and considerable price reductions could reasonably be expected.

(10) CONCLUSIONS

It has been possible to show both theoretically and practically that the non-integral index structure outlined in the paper can give a satisfactory index signal, and that it has several advantages over the only known practical alternative.[18] The advantages are:

(*a*) The tube construction is simpler since only one anode button is required and there is no collector-electrode or screen insulation problem. Also, the gun is less complicated.

(*b*) It is much easier to make a phosphor screen with a consistent performance than a secondary-emission screen.

(*c*) There is no transit-time variation, and no problems due to the different relative positions in the horizontal direction of the writing and pilot beams when they are deflected. The index structure can therefore be parallel to the colour-screen structure. This eases the registration problem in the screen manufacture.

(*d*) It is easier to use the phase-compensating arrangement of Fig. 11.

(*e*) The use of a high-frequency carrier in the secondary-emission system results in low-level index signals which are prone to outside interference. In the photo-electric system, all electric index signals are at a comparatively high level and do not suffer from this fault.

The disadvantages are:

(*a*) A photo-multiplier is needed, the cost of which is at present unknown.

(*b*) The index-signal limiting problem is much more severe.

(*c*) The index signal is less positive in nature since, by itself, it is ambiguous during the horizontal scan and relies on further information which is present only at the beginning of the scan.

On the practical side, work in the laboratory has so far been limited to the use of 14 in rectangular tubes of 70° deflection angle, with a screen composed of about 350 phosphor triads spaced 32 to the inch. The highlight brightness was 15–20 ft-lamberts when the tube was operated at 25 kV and a mean beam current of about 250 μA. Under these conditions, life tests have shown that the tubes have adequate life. The circuits were designed primarily to give maximum performance and flexibility and to investigate various aspects of this and other systems.

No attempt has been made to produce anything approaching a commercial design of receiver, and at this early stage it is not

possible to say whether an index display would meet the main requirements of such receivers. These are:

(i) A good-quality colour picture of adequate brightness.
(ii) Reliability and ease of adjustment.
(iii) A reasonable price.
(iv) The ability to produce a good black-and-white picture.

It is, however, possible to make some general comments and comparisons with other types of tube.

Although it has been shown that it is quite possible to convert the N.T.S.C. signal to suit the needs of an index tube, nevertheless some minor modifications to the transmitted signal might simplify the signal processing if heterodyne techniques are used. These modifications are unlikely to affect the use of other types of display.

The parameters in an index tube which affect the quality of the colour picture are many, and include the number of triads per inch, the size of tube, brightness, deflection angle, contrast ratio, the degree of ambient light protection, the use of lenticular filters,[20] e.h.t. regulation, photo-cathode area, the quality of deflector coils, cathode-loading, transfer characteristics of the signal circuits, etc., and several of these factors are interrelated.

There is a great deal of development required in this field before it is possible to say that a satisfactory compromise would be commercially acceptable.

On the question of reliability, this can only be assessed when a commercial prototype receiver has been designed, but it is a basic philosophy of the index system that much of the complexity has been taken out of the tube and put into the circuits. This makes for great flexibility of design and ease of adjustment. It could be argued, however, that circuits employing active devices such as valves will almost certainly be less reliable in operation than the static auxiliary components associated with a shadow-mask tube.

One definite feature of the index tube is its ability to produce a good black-and-white picture with no circuit complexity, i.e. the tube is treated exactly as a normal monochrome tube. However, this does mean that the 'whiteness' must be built into the screen when the tube is initially made, and the use of a single gun means that one of the advantages of a three-gun tube is lost. This is the ability of being able to adjust the three beam currents to take into account any variations in the relative efficiency of the three phosphors, which may occur for a variety of reasons.

Although the tolerances in screen manufacture are extremely close and possibly more difficult to achieve than in other types of tube, the advent of new techniques such as frit sealing, and the advantage of having no mechanical structure to align with the screen, open up new possibilities for phosphor deposition. This and allied subjects are being dealt with in the companion paper already referred to.

(11) ACKNOWLEDGMENTS

This work was carried out on behalf of Thorn Electrical Industries, Ltd. and Sylvania Electric Products, Inc. The authors would like to thank their colleagues in these two organizations for their assistance, and in particular, Mr. A. E. T. Brown and his team for providing the many varieties of experimental colour tubes which were required for this investigation.

(12) REFERENCES

(1) Fleming-Williams, B. C.: British Patent No. 793 351, 1954.
(2) Lindsay, H. R., Oxenham, J. K., and Fleming-Williams, B. C.: British Patent No. 793 352, 1954.
(3) 'General description of Receivers for the Dot-Sequential Color Television System which Employ Direct-View Tri-Color Kinescopes', *RCA Review*, 1950, **11**, pp. 228.
(4) Law, H. B.: 'A Three-Gun Shadow-Mask Color Kinescope', *Proceedings of the Institute of Radio Engineers*, 1951, **39**, p. 1186.
(5) Dressler, R.: 'The PDF Chromatron—a Single or Multi-gun Tri-Color Cathode-Ray Tube', *ibid.*, 1953, **41**, p. 851.
(6) Muller, E. N.: United States Patent No. 2630548, 1937 (Luxemburg).
(7) Stevens, W. H.: British Patent No. 603080, 1944.
(8) Clapp, R. G., Creamer, E. M., Moulton, S. W., Partin, M. E., and Bryan, J. S.: 'A New Beam-Indexing Color Television Display System', *Proceedings of the Institute of Radio Engineers*, 1956, **44**, p. 1108.
(9) Barnett, G. F., Bingley, F. J., Parsons, S. L., Pratt, G. W., and Sadowsky, M.: 'A Beam-Indexing Color Picture Tube—The Apple Tube', *ibid.*, 1956, **44**, p. 1115.
(10) Bloomsburgh, R. A., Boothroyd, W. P., Fedde, G. A., and Moore, R. C.: 'Current Status of Apple Receiver Circuits and Components', *ibid.*, 1956, **44**, p. 1120.
(11) Philco Corporation: British Patent No. 753441, 1950.
(12) Philco Corporation: British Patents Nos. 753442 and 753555, 1950.
(13) Philco Corporation: British Patent No. 753451, 1951.
(14) Bowie, R. M.: United States Patent No. 2827591, 1954.
(15) Chatten, J. B., and Gardner, R. A.: 'Accuracy of Colour Reproduction in the Apple System', *Institute of Radio Engineers National Convention Record*, 1957, **5**, Part 3, p. 230.
(16) Bond, D. S., Nicoll, F. H., and Moore, D. G.: 'Development and Operation of a Line-Screen Color Kinescope'. *Proceedings of The Institute of Radio Engineers*, 1951, **39**, p. 1218.
(17) Oxenham, J. K., and Graham, R.: British Patent Applications Nos. 31902 and 39233, 1957.
(18) Payne, D., Colgate, H. R., Moulton, S. W., Comeau, C. P., Kelley, D. P.: 'Recent Improvements in the Apple Beam-Indexing Color Tube', *Institute of Radio Engineers National Convention Record*, 1957, **5**, Part 3, p. 238.
(19) McIlwain, K., and Dean, C. E. (Eds.): 'Principles of Colour Television' (Chapman and Hall, 1956).
(20) Bloomsburgh, R. A., Hopengarten, A., Moore, R. C., and Wilson, H. H.: 'An Advanced Color Television Receiver using a Beam Indexing Picture Tube'. *Institute of Radio Engineers National Convention Record*, 1957, **5**, Part 3, p. 243.

(13) APPENDICES

(13.1) The Calculation of Cross-Modulation

The drive applied to the grid of the cathode-ray tube is $V_s = V_m + V_c \cos(\omega_s t - \phi)$, and the beam current will therefore be $I_a = [V_m + V_c \cos(\omega_s t - \phi)]^\gamma$. By Fourier analysis this waveform can be expressed as

$$I_a = \frac{A_0}{2} + \sum_{k=1}^{\infty} A_k \cos k(\omega_s t - \phi) \quad . \quad . \quad . \quad (6)$$

where the coefficients A_0 and A_k are functions of V_m, V_c and γ. The index signal I can also be expressed as a Fourier series:

$$I = I_a\left[\frac{B_0}{2} + \sum_{l=0}^{\infty} B_l \cos l\left(\frac{m}{n}\omega_s t\right)\right] \quad . \quad . \quad . \quad (7)$$

where the coefficients B_0 and B_l are functions of the index stripe width and the spot size and shape.

Substituting eqn. (6) into eqn. (7),

$$i = \left[\frac{A_0}{2} + \sum_{k=1}^{\infty} A_k \cos k(\omega_s t - \phi)\right]\left[\frac{B_0}{2} + \sum_{l=1}^{\infty} B_l \cos l\left(\frac{m}{n}\omega_s t\right)\right] \quad . \quad . \quad . \quad . \quad (8)$$

The component of this signal at index frequency is

$$I_{m/n\omega_s} = \tfrac{1}{2}A_0B_1 \cos\frac{m}{n}\omega_s t + \tfrac{1}{2}\sum_{r=1}^{\infty} A_{mr}B_{nr-1} \cos\left(\frac{m}{n}\omega_s t - mr\phi\right) + \tfrac{1}{2}\sum_{r=1}^{\infty} A_{mr}B_{nr+1} \cos\left(\frac{m}{n}\omega_s t + mr\phi\right)$$

$$= D \cos\frac{m}{n}(\omega_s t - \theta) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

where $$\theta = \frac{n}{m} \arctan\left[\frac{\sum_{r=1}^{\infty} A_{mr}(B_{nr-1} - B_{nr+1}) \sin mr\phi}{A_0B_1 + \sum_{r=1}^{\infty} A_{mr}(B_{nr-1} + B_{nr+1}) \cos mr\phi}\right] \quad . \quad . \quad . \quad . \quad (10)$$

This angle θ represents the phase error of the index signal after it has been frequency-changed back to writing frequency.

Substituting $m = 3$ into eqn. (10) and ignoring terms in which $r>1$, since A_{3r} is negligible when $r>1$,

$$\theta = \frac{n}{3} \arctan\frac{A_3(B_{n-1} - B_{n+1}) \sin 3\phi}{A_0B_1 + A_3(B_{n-1} + B_{n+1}) \cos 3\phi} \quad (11)$$

By differentiating eqn. (11) and equating to zero, the value of ϕ which maximizes θ can be found, and when this is substituted back into eqn. (11) we get

$$\theta_{max} = \frac{n}{3} \arctan\frac{A_3(B_{n-1} - B_{n+1})}{\sqrt{[(A_0B_1)^2 - A_3^2(B_{n-1} + B_{n+1})^2]}} \quad (12)$$

Fig. 2 shows θ_{max} plotted against d/p. Eqn. (9) shows that the index signal $I_{(m/n\omega_s)}$ is comprised of two parts, the wanted signal

$$\tfrac{1}{2}A_0B_1 \cos\frac{3}{n}\omega_s t$$

and the unwanted signal

$$\frac{A_3}{2}\left[B_{n-1} \cos\left(\frac{3}{n}\omega_s t - 3\phi\right) + B_{n+1} \cos\left(\frac{3}{n}\omega_s t + 3\phi\right)\right]$$

This index signal becomes unusuable if either the wanted signal is zero or the unwanted signal is larger in amplitude than the wanted signal. In the latter case, the index signal becomes completely dependent in phase upon the phase of the modulating signal, and the system is unstable.

For $m/n = 3$, the wanted signal is

$$\tfrac{1}{2}A_0B_1 \cos 3\omega_s t$$

and the unwanted signal is

$$\frac{A_3}{2}[B_0 \cos(3\omega_s t - 3\phi) + B_2 \cos(3\omega_s t + 3\phi)].$$

As the spot size increases, B_1 gets smaller, but the A_3B_0 term of the unwanted signal stays constant since B_0 is independent of spot size. The condition of instability is reached when these two signals are equal in amplitude, i.e. when $A_3(B_0 + B_2) = A_0B_1$, and this occurs when $d/p = 0{\cdot}22$ and $\theta_{max} = 30°$. This is shown as point A on the graph.

For $m/n = \frac{3}{2}$ or $\frac{3}{4}$, the wanted signals are

$$\tfrac{1}{2}A_0B_1 \cos\tfrac{3}{2}\omega_s t$$

and

$$\tfrac{1}{2}A_0B_1 \cos\tfrac{3}{4}\omega_s t$$

respectively, and the unwanted signals,

$$\frac{A_3}{2}[B_1 \cos(\tfrac{3}{2}\omega_s t - 3\phi) + B_3 \cos(\tfrac{3}{2}\omega_s t + 3\phi)]$$

and

$$\frac{A_3}{2}[B_3 \cos(\tfrac{3}{4}\omega_s t - 3\phi) + B_5 \cos(\tfrac{3}{4}\omega_s t + 3\phi)]$$

respectively. As the spot size increases, all the B coefficients decrease, and the index signal is unusable when the wanted signal is zero, i.e. when $B_1 = 0$. It can be shown that this occurs when $m/n \times d/p = 1$, i.e. when $d/p = n/m$. These two points are also shown as points A on the graphs.

(13.2) The Effect of the Loop Time Delay

It is assumed that, within the frequency band considered, the phase/frequency characteristic of the circuits is linear, i.e. $d\phi/d\omega = \text{constant} = T$.

Let the index frequency ω_i be given by $\omega_i = d\phi/dt$, where

$$\phi = \tfrac{3}{2}\left[\omega_s t + \mathrm{F}(t)\right] \quad . \quad . \quad . \quad . \quad (13)$$

where ω_s is a constant and $\mathrm{F}(t)$ is a function of time and represents a phase perturbation due to a variation of scan velocity.

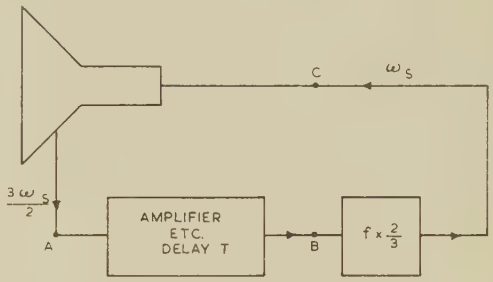


Fig. 10.—Simple index loop.

Then the index signal phase at A (Fig. 10) is given by

$$\phi_a = \tfrac{3}{2}[(\omega_s t + \mathrm{F}(t)] \quad . \quad . \quad . \quad . \quad (14)$$

and the phase at B by

$$\phi_B = \tfrac{3}{2}[\omega_s(t - T) + \mathrm{F}(t - T)] \quad . \quad . \quad (15)$$

The writing-signal phase at C is given by

$$\phi_c = \tfrac{2}{3}\phi_B = \omega_s(t - T) + \mathrm{F}(t - T) \quad . \quad . \quad (16)$$

assuming that the act of multiplication by $\frac{2}{3}$ gives no further phase delay.

The writing-signal phase error, ϕ_e, is then given by

$$\phi_e = \tfrac{2}{3}\phi_a - \phi_c = \omega_s T + \mathrm{F}(t) - \mathrm{F}(t - T) \quad . \quad (17)$$

$\omega_s T$ represents a constant phase error and can be removed by a simple phase shift.

Expanding by Taylor's series,

$$\phi_e = T\mathrm{F}'(t) - \frac{T^2}{2!}\mathrm{F}''(t) + \frac{T^3}{3!}\mathrm{F}'''(t) \quad . \quad . \quad (18)$$

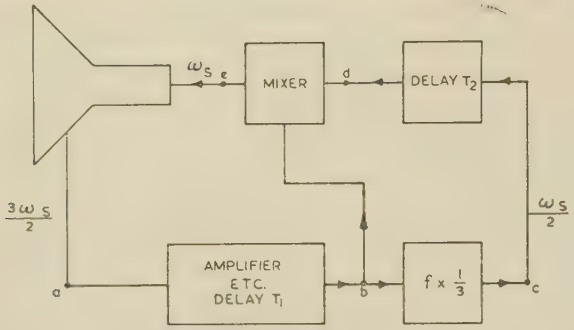


Fig. 11.—Modified index loop.

Using the same techniques for Fig. 11,

$$\phi_a = \tfrac{3}{2}[\omega_s t + \mathrm{F}(t)] \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

$$\phi_b = \tfrac{3}{2}[\omega_s(t - T_1) + \mathrm{F}(t - T_1)] \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

$$\phi_c = \tfrac{1}{3}\phi_b = \tfrac{1}{2}[\omega_s(t - T_1) + \mathrm{F}(t - T_1)] \quad . \quad . \quad . \quad . \quad . \quad (21)$$

$$\phi_d = \tfrac{1}{2}[\omega_s(t - T_1 - T_2) + \mathrm{F}(t - T_1 - T_2)] \quad . \quad . \quad . \quad . \quad (22)$$

$$\phi_e = \phi_b - \phi_d$$

$$= \omega_s t + \frac{\omega_s}{2}(T_2 - 2T_1) + \tfrac{3}{2}\mathrm{F}(t - T_1) - \tfrac{1}{2}\mathrm{F}(t_1 - T_1 - T_2) \quad (23)$$

The phase error is given by

$$\phi_f = \tfrac{2}{3}\phi_a - \phi_e$$

$$= -\frac{\omega_s}{2}(T_2 - 2T_1) + \mathrm{F}(t) - \tfrac{3}{2}\mathrm{F}(t - T_1) + \tfrac{1}{2}\mathrm{F}(t - T_1 - T_2) \quad (24)$$

Let $T_2 = 2T_1$. Then

$$\phi_f = \mathrm{F}(t) - \tfrac{3}{2}\mathrm{F}(t - T_1) + \tfrac{1}{2}\mathrm{F}(t - 3T_1) \quad . \quad . \quad (25)$$

and expanding by Taylor's series,

$$\phi_f = \frac{3T_1^2}{2!}\mathrm{F}''(t) - \frac{8T_1^3}{3!}\mathrm{F}'''(t) + \ldots \quad . \quad . \quad (26)$$

(13.3) Demonstration

The system demonstrated at The Institution on the 1st May, 1961, differed in some respect from that described in the paper. The main differences lie in the tube itself, which was a 21 in frit-sealed tube with the following specification:

Screen area	14½ in × 11½ in
Screen efficiency	50 ft-L/mA
Pitch of colour triads ..	25·3 triads/in = 0·039 5 in pitch
Phosphor stripe width ..	0·006 6 in
Number of phosphor triads	367
Triode gun with:	
Nominal cut-off ..	−110 V at 25 kV
Drive for 1 mA ..	60 V
Spot size	0·009 in width at 50% brightness level up to 2 mA; 0·012 in height at 100 μA; 0·050 in height at 1·5 mA

DISCUSSION BEFORE THE ELECTRONICS AND COMMUNICATIONS SECTION, 1ST MAY, 1961

Dr. R. D. A. Maurice: The television world has been awaiting the completion of the development of a beam-index colour tube for many years and the authors have taken a major step forward. I refer to their use of an index structure having a periodicity which is different from that of the colour-phosphor stripes. The difficulties which they have overcome and those which I suspect still remain are formidable.

I took the trouble to apply eqns. (1), with the results shown in Fig. A. Evidently Fig. 1 has not been drawn to scale and in Fig. A it can be seen how the negative primaries, which

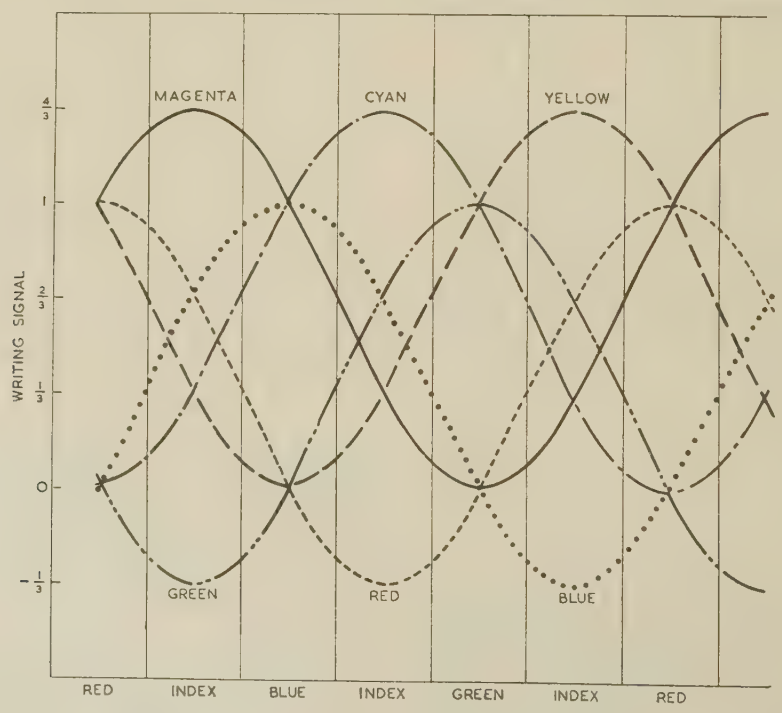


Fig. A.—Writing signals for the six primary colours.

require the value unity at the two appropriate positive primaries, force the writing beam to exceed unity at the colour occurring between the two required positive primaries.

Turning to Section 6, one can realize that the signal/noise ratio of the photo-indexing signal must needs be high since an observable degradation of the picture can be caused by a r.m.s. deviation of horizontal position of as little as one-fifth of a picture element.

It seems that the monochrome signal M is composed partly of the full-bandwidth luminance signal Y and partly of band-width-restricted chrominance signals. Would I be right, therefore, in saying that, quite apart from questions of failure of constant luminance due to the use of gamma-corrected primary signals, the monochrome signal will not have the horizontal resolution contained within it that the normal luminance signal Y does, in fact, carry? If this is so, the sharpness of pictures containing some colour will be less on a beam-index tube than with a 3-gun tube.

The paper indirectly raises the question whether the transmitted signals should be tailored to suit a beam-index tube or a 3-gun tube; I would say that, until we have had as much experience with a beam-index tube as we have had with the 3-gun tube, we ought to leave the N.T.S.C. signal as it is at present. This could mean that, if a beam-index tube does not come on the market until after colour transmissions in Europe have started, the beam-index receiver will be at the slight disadvantage of having to use more complicated circuits than would have been the case otherwise.

Dr. F. E. Jones: A report by the Department of Scientific and Industrial Research is quoted as saying that in this country 12% of the output of the electronics industry is spent in research, and that it is more than in any other industry except the aeroplane industry. It also says that 7% of the scientists and engineers employed in industry work in the electronics industry. In the United States the percentage figures are even higher. This indicates a large amount of scientific and engineering talent.

The original publications on the shadow-mask tube appeared more than 10 years ago, and presumably work on it was done even earlier. It is fairly well accepted that this is the only display tube available in manufacture for colour-television receivers now, and even it will be associated, to put it mildly, with a great deal of hesitancy on the part of people in accepting it in a system.

A colour display and system are desirable in television, and possibly also in a certain number of industrial applications, and the statistics which I have given relating to effort and time give some indication of the difficulties with which the team interested in a colour television are confronted. In spite of the efforts of the present team, a tremendous amount of work and money must be put into the system before it becomes a practical proposition. Those of us who have some responsibility in making decisions about colour are very anxious to get assessments on what future work is needed and what investment in these fields will bring. As in black-and-white picture systems, and still more in colour, television is a compromise between quality and cost.

The points I wish to raise are in this area. I believe that the original beam-index tube used a spacing of 16 triads/in. Most people felt that the lines were too apparent in its structure. The present tube uses 32 triads/in, but the brightness comes down to 15–20 ft-L compared with 60 ft-L for the original tube and 30–40 ft-L of the presently available shadow-mask tube. Will the authors comment on this compromise between the triad spacing and the brightness?

How does the beam width in the authors' tube compare with black-and-white tube practice? It seems to me that some beam trimming may have been used.

It does not seem necessary to use a circular spot in this application, but I gather that there are some curves available to show what is the effect of an elongated spot.

Finally, from the practical side, television customers, if there are any these days—and there seems to be some doubt about it—are accustomed to slim-line sets. What difficulties are foreseen in coming down from the present 70° to something wider, which would perhaps have a modern appeal?

Dr. E. L. C. White: I have always thought that some kind of single-gun tube will be the final answer to the colour-receiver problem. In addition to the disadvantages of the 3-gun tube mentioned in the paper, there is also the low efficiency of the shadow-mask itself, which allows about 15% of the total beam current to pass through.

Single-gun tubes have been tried in several forms without as yet being so obviously the right way to do it as to challenge the 3-gun tube. First there was a single-gun shadow-mask tube in which the one beam was bent so as to appear to come from different places in succession, simulating the 3-gun shadow-mask tube. This gets over the problem of matching gun characteristics but leaves the registration problem. Secondly, there is the Chromatron, which the authors described. Then there is the Apple tube, and fourthly, there is its close relation, about which we have just heard. But in all cases the method of working has been to direct the beam in a regular sequence to the three phosphor elements which form effectively one picture element, and to modulate the beam in dependence on this phosphor-selection sequence in combination with the luminance and chrominance signals. The overall result is similar to the one obtained with a 3-gun tube, and this is fairly well suited to the N.T.S.C. type of signal. It is worth noting that all the one-gun types, except perhaps the Chromatron, can be operated in another way, in which luminance signals only are used to modulate the beam, and chromaticity instead of chrominance signals are used to steer it to such a position or in such a pattern as to excite the phosphors in the correct ratios to give the colour required. I believe that such a system is worth a trial, but as chromaticity signals would not suit the 3-gun tube, the system would have to show marked superiority to stand any chance of acceptance. Its advantages would be:

(*a*) The luminance signal is correct and would therefore give better compatible black-and-white pictures.

(*b*) It could be generated with full definition from a single camera tube without loss of luminance definition due to camera registration problems.

(*c*) Differences of gamma at the camera and receiver do not alter hue or saturation but alter only brightness.

Will the authors comment on these possibilities?

Mr. J. Sharpe: The indexing tube described by the authors requires a photomultiplier tube having good limiting characteristics and also capable of being produced relatively cheaply. In the paper, it is suggested that a gain of only 5000 is needed and that this would simplify the tube construction and so aid in reducing cost.

I should like to point out that the number of dynodes in excess of three or four does not make an enormous difference to the cost of a photo-multiplier tube in the quantities which would be needed for this application and that there is a positive advantage in having as high gain as possible in order to improve the characteristics of the tube as a limiter.

Tests on one of our 11-stage photomultiplier tubes type 9524 B have shown that by operating with the anode at almost the same potential as the last dynode, an output current of 100 μA was obtainable for an input cathode current of 0·02 μA and varied by less than 2 to 1 as the cathode current varied from 0·002 to 0·6 μA. This extremely effective limitation, obtained at the cost of greater complexity in the photomultiplier tube, should make

the circuits in the receiver appreciably simpler than would be the case if a lower-gain tube were used.

Mr. A. V. Lord: The authors have discussed the beam-index tube as used with the N.T.S.C. system. Will they comment upon its suitability for use with the S.E.C.A.M. system? Complete demodulation of the composite colour signal would, presumably, be necessary before forming the colour-writing signal.

The paper draws attention to the need for a very linear line-scan when using a beam-index tube and describes suitable but somewhat inefficient circuit arrangements. At present only the high-quality monitor may be expected to achieve a linearity within 0·5% and it is quite usual to find errors of 5% in domestic receivers. Are the authors confident that further development can produce an efficient line-scan arrangement with the degree of linearity they require?

Finally, the paper describes a 14in tube having 32 triads/in and the lecture has revealed a 21in tube with 25 triads/in. In view of the fact that even the 405-line system has 490 picture elements along each line, are the authors satisfied that either tube will have adequate resolution?

Mr. D. A. Rudd: In discussing the factors affecting the choice of indexing frequency the authors state in Section 8 that there is an improved light yield from the indexing phosphors at an angular frequency of $3\omega_s/4$ compared with $3\omega_s/2$. Any such improvement will, I presume, be due to the reduced effect of the phosphor decay time at the lower frequency. However, reduction of the indexing frequency reduces the comparative width of the indexing stripe and therefore reduces the fundamental component of the indexing signal. Thus, Fig. 4 shows that the ratio of index-stripe width to index-stripe separation is 1 : 8 for the $3\omega_s/4$ system, whereas this ratio will be 1 : 4 for the $3\omega_s/2$ system. Taking into account the finite spot width, it would appear that the advantage gained by the reduced effect of the phosphor decay time when indexing at $3\omega_s/4$ may be almost completely eliminated by the corresponding reduction in the fundamental component of the screen structure. Will the authors comment on this point?

Two methods of using the photomultiplier to limit the amplitude of the indexing signal are described, one by operating the photomultiplier in a non-linear manner, and one by modulating the dynodes. It appears to me that the use of either of these methods would cause electron-transit-time variations within the photomultiplier itself and thereby introduce indexing phase errors. These transit-time variations would have to be less than 10^{-9}sec in order to have an inappreciable effect and I wonder if, in fact, they have been observed by the authors.

Dr. A. J. Biggs: In order to maintain the sensing signal the beam current must never fall below a certain value. Presumably this can be covered by automatic means, but does this occasion any additional problem in control?

The life of this tube will be very closely related to the triad structure. The finer the spot, the bigger the problems in getting good life. What is the order of this effect? How does the life of the tube with the kind of structure described compare with the life of the 3-gun shadow-mask tube?

If we go on to higher-definition systems we are asking for smaller spots and more problems in the sensing tube. If the authors believe, as I do, that this is a most promising technique for a colour display, does it mean that in the future we shall be interested in keeping the definition at just a reasonable level and that we shall not ask for higher definition in order to prevent our problems in colour performance from becoming greater?

Our present trend is to cut out the horizontal lines and to substitute vertical lines. If we are sufficiently enthusiastic about this sensing type of display, does it mean that we should be in favour of sticking to 405 lines because you have to sit a little way back to avoid seeing the scanning lines and then you will not see the triad lines either?

Mr. L. C. Jesty: First, I should like to try and clarify the position which this paper has with regard to other work in the same field, particularly in the United States. Index systems have been under active investigation in this country for a long time, certainly since 1949. The present investigation began and the main objectives were determined some time before the first release in the United States of a twin-beam secondary-emission index tube and receiver in 1956.* The authors' work has developed along other and perhaps more elegant lines, using a truly single-beam tube with a simpler screen structure. The first demonstrations of this were given privately in October, 1958, using a 14in rectangular tube and a $\frac{3}{4}$ index sequence. This stage was arrived at without reference to any further work in the United States following the 1956 release. Later it came to our knowledge that there was development there along similar lines.† During the last two years, we have had a number of very useful exchanges of opinion with American engineers and others. We feel very strongly that collaboration is essential if a colour-television receiver such as that described is to achieve satisfactory commercial realization.

My second point concerns the shadow-mask tube. This is undoubtedly well established but it needs some healthy competition. This would provide the incentive both in this country and the United States for more rapid progress than hitherto in the commercialization of colour television.

The single-beam photo-index system described by the authors is running about 10 years behind the shadow-mask tube at the moment. Some people think it is too late to try and compete with the latter, and that in another 10 years an entirely new type of colour display may evolve, which will revolutionize the whole position. A decision has to be made, therefore, with the type of situation shown in Fig. B in mind. If the shadow-mask tube is reaching the limit of development after 10 years, there is a better chance that a competitor might catch and even surpass it in about 3 years, along the lines of curves A_I and A_S. Should the shadow-mask continue to improve, however, competitors might never catch up. The worst case is illustrated

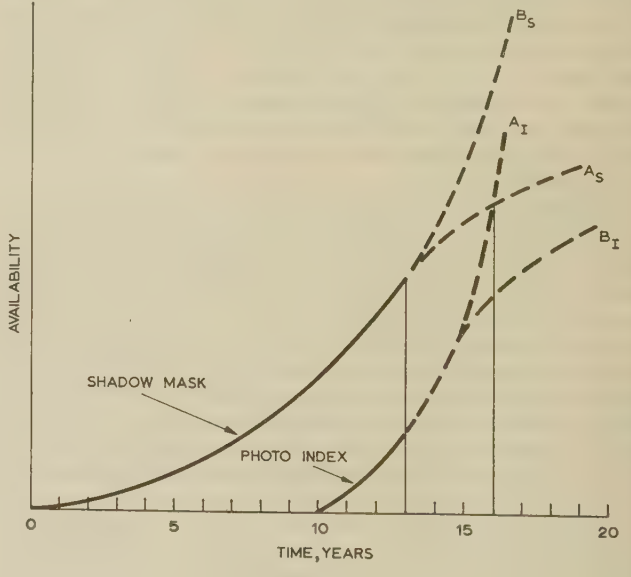


Fig. B

* Reference 8 in the paper.
† See, for example U.S. Patent No. 2892123, June, 1959.

by curves B_I and B_S, which condition would not justify continuation with an alternative. I feel that curves A are nearer the truth. This is the situation which has to be evaluated now, when in fact there is a great desire in certain quarters to introduce a colour-television service into this country. To do so would obviously make it unattractive to invest in a potential alternative colour tube.

Thirdly, I should like to emphasize that a lot of the enthusiasm for the shadow-mask tube comes from circuit engineers who have had no alternative to practise their skill on. They do not have to make the tube but only to use it. We need a simple colour tube and the rest will follow.

Finally, I would like to name this tube. We already have the Apple tube, and the Banana will be shown shortly. I propose to call this the Zebra tube. It has vertical stripes; there is a degree of finality about Z; and in this country, Zebras enable people to negotiate a dangerous path in comparative safety.

Mr. W. H. Buchanan: The major contribution which new tubes and new techniques can make to colour television lies in cost. In this connection the tube described will be easier to maintain by the user than the shadow-mask tube, in that it appears to be relatively easy to re-gun it. Will the authors comment?

Does the thickness of aluminium to shield the main light from the photocell differ in any great respect from the optimum aluminium thickness appropriate to the 25 kV electron beam?

THE AUTHORS' REPLY TO THE ABOVE DISCUSSION

Messrs. R. Graham, J. W. H. Justice and **J. K. Oxenham** (*in reply*): In reply to Dr. Maurice, the full-bandwidth *Y*-signal is applied to the grid of the cathode-ray tube, as in common practice in other types of colour display. In addition, three narrow-bandwidth colour signals are added to the luminance signal, namely the $M-Y$ signal, the diode correction signal, and the chroma signal on a carrier of angular frequency ω_s (this compares with the three narrow-bandwidth $R-Y$, $G-Y$, $B-Y$ signals added to the luminance signal in a 3-gun tube). The final result is that the transient response is very similar on a beam-index tube to that on a 3-gun shadow-mask tube, neglecting any effects due to the finite size or coarseness of the phosphor structure.

Dr. Jones quotes figures for the 14in tube described in the paper. The 21in tube demonstrated has somewhat different parameters and these have been tabulated in Section 13.3 in the paper, from information supplied by Mr. A. E. T. Brown, Mr. R. C. Coleclough and Mr. D. E. Ridout. We have done no work on determining the optimum compromise between brightness, triad spacing and deflection angle, and the figures quoted merely represent a typical set of conditions which have been achieved. The relationship between these three important factors is complex, and depends also upon a number of other parameters. It does, however, seem realistic to increase the deflection angle to 90°, although some decrease in brightness and increase in triad spacing would probably be necessary. The diameter of the beam in the deflector coils is 0·18in, and no beam trimming is employed.

Dr. White describes a system in which a luminance signal and a chromaticity, rather than a chrominance, signal are used. This system was envisaged early on (see Reference 1 in the paper) but was abandoned since the straightforward grid-modulation approach is more suitable for a N.T.S.C.-type signal. It has the disadvantage that an *M*-signal is still required to modulate the tube, and this signal can no longer be derived easily from the transmitted information. The cross-modulation characteristic of this method of modulation has not been analysed, but it seems reasonable to assume that it would be very similar to the system described in the paper.

We have used the photomultiplier type 9524B described by Mr. Sharpe with success. Care must be exercised when using the photomultiplier in a non-linear mode since it can itself cause additional cross-modulation due to the $\omega_{s/2}$ and ω_s signals present in the input.

Mr. Lord is correct in his assumption that a demodulation-remodulation technique would be necessary for use with the S.E.C.A.M. system. We do not envisage that the horizontal-linearity problem will be a particularly serious one, especially as it would be possible for a commercial receiver to tailor the triad spacing to fit the naturally exponential scan. On the question of horizontal definition, it is worth while pointing out that the tube demonstrated has a total of 1100 vertical phosphor stripes in the width of the picture, but so far no subjective tests have been carried out to determine to what extent the phosphor structure degrades the picture definition.

It is presumed that Mr. Rudd's query concerns the signal/noise ratio of the index signals in the two systems. When all effects have been taken into account, the signal/noise ratio of the $3\omega_s/4$ system is about twice that of the $3\omega_s/2$ system. No effects have so far been observed which are attributable to finite transit time in the photomultiplier.

On the question of tube life, as posed by Dr. Biggs, we have no information on the shadow-mask tube, but our own life tests show that the spot size is still adequate for the requirements of the system after 2000h with a modulated signal and 500μA mean current.

In reply to Mr. Buchanan, provided that the gun is sealed in at the correct orientation, the tube can be re-gunned using the same techniques as are used for monochrome tubes. The thickness of aluminium used is chosen to be appropriate to the final accelerating voltage of 25kV, and under these conditions very little light from the colour phosphors reaches the photomultiplier.

621.374.44 : 621.391 : 621—52

The Institution of Electrical Engineers
Paper No. 3612 M
Sept. 1961

A RANDOM PULSE GENERATOR WITH VARIABLE MEAN RATE

By J. L. DOUCE, M.Sc., Ph.D., Graduate, and B. G. LEARY, B.E., Ph.D., Associate Member.

(*The paper was first received* 20*th October*, 1960, *and in revised form* 27*th February*, 1961.)

SUMMARY

The paper describes the methods by which a random pulse train with predetermined statistical properties may be produced. It is required that the mean pulse rate shall be variable over a wide range and that the probability of a pulse occurring at any instant shall be independent of the previous history of the pulse train. This second condition is satisfied if the time of occurrence of the pulses obeys the Poisson distribution, and several methods of realizing this distribution are investigated.

The most satisfactory practical means of obtaining the required pulse train is found to be the sampling of a random signal, an output pulse being given whenever the random signal exceeds a predetermined amplitude. The theoretical requirements are discussed and the circuit techniques presented. Automatic control of mean rate is incorporated in the device described.

(1) INTRODUCTION

The Poisson distribution specifies the probability that a definite number of events, n, will be observed in a time interval T when the events are occurring at a mean rate ν events per unit time. To satisfy this distribution the events must comply with certain conditions, which may be summarized as follows:

(*a*) The occurrence of an event at any particular time is independent of the previous history.
(*b*) As the time of observation, δT, approaches zero, the probability of observing one event tends to $\nu\delta T$, and the probability of observing more than one event tends to zero more rapidly than δT.

The probability of finding zero pulses in a time T, P(0, T), is readily derived with the aid of the above conditions. Consider the time interval T divided into small intervals δT. In any such interval it is certain that some number of pulses will occur, i.e. $1 = \mathrm{P}(0, \delta T) + \mathrm{P}(2, \delta T) + \ldots$. Moreover, as $T \to 0$, $\mathrm{P}(1, \delta T) \to \nu\delta T$, and P(2, δT), etc., are very much less than P(1, δT) by requirement (*b*). Hence $\mathrm{P}(0, \delta T) \to 1 - \nu\delta T$. For zero pulses in time T there must be no pulses in every interval δT comprising time T. Since the occurrence of a pulse at any time is independent of the instant of observation, the overall probability of zero pulses in time T is

$$P(0, T) = [P(0, \delta T)]^{T/\delta T} = (1 - \nu\delta T)^{T/\delta T}$$

As $\delta T \to 0$, this gives $\mathrm{P}(0, T) = \varepsilon^{-\nu T}$.

This expression is most useful for proving the distribution of a pulse sequence. The time intervals between pulses are determined and a frequency histogram is plotted and compared with the theoretical exponential function.

Fig. 1 shows the probability of observing various numbers of events as a function of time of observation, νT. The derivation of the higher functions is a more lengthy procedure and is not presented here. The complete expression is[1]

$$\mathrm{P}(n, T) = \frac{(\nu T)^n \varepsilon^{-\nu T}}{n!} \quad . \quad . \quad . \quad . \quad . \quad (1)$$

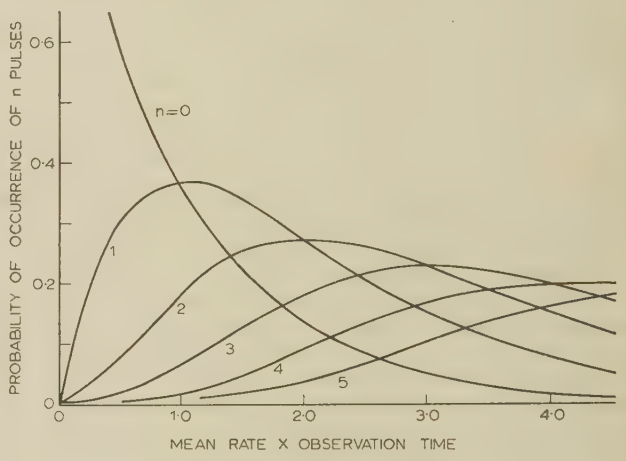


Fig. 1.—Probability of occurrence of n events.

Many events of practical importance obey the Poisson distribution, e.g. the arrival of telephone calls at any exchange and the output of several identical batch production units, and hence the need arises for a pulse generator of this form in computer analysis of queueing and storage problems.

(2) PREVIOUS DEVELOPMENTS

The decay of a radioactive substance accompanied by the emission of α-particles is a random process,[8, 9] and from the measurements of Rutherford and Geiger[2] the time occurrence of these events has been shown to have a Poisson distribution. This phenomenon is not favoured for general use, however, in view of the hazards present and the practical difficulty in varying the mean pulse rate. Similarly, mechanical and microscopic phenomena (e.g. thermionic emission) are to be avoided.

A considerable volume of work has been published on the statistical properties of a continuously varying random signal, and two lines of approach suggest that the zero crossings of such a signal obey the Poisson event distribution of eqn. (1). Before discussing the practical results it is convenient to summarize some previous analysis.

Involved theoretical work by Bendat[3] shows that the Poisson distribution is a possible solution to the equations derived for the time distribution of the zero crossings of a Gaussian signal, and a more direct analysis derives the autocorrelation function, and hence the power spectrum, of a square wave whose zero crossings are specified to be of the desired form. A Gaussian random signal, passed through appropriate frequency-dependent elements and applied to a limiter, should then give the appropriate waveform. Considerable effort has been expended on the production of this waveform, and the theoretical basis of the work is as follows.

A changing-polarity square wave whose switching times obey a Poisson distribution has an autocorrelation function given by[4]

$$\phi(\tau) = \beta^2 \exp - 2\nu|\tau| \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Dr. Douce and Dr. Leary are in the Electrical Engineering Department, Queen's University, Belfast.

and hence a power spectrum of the form

$$G(\omega) = 2\beta^2 T \frac{1}{1 + (\omega T)^2}$$

where β is the amplitude of the square wave and T is given by $1/2\nu$.

When a large-amplitude Gaussian signal is passed through a symmetrical limiter, the autocorrelation function of the output, $\phi_0(\tau)$, is related to that of the input, $\phi_i(\tau)$, by the well-known expression[5]

$$\frac{\phi_0(\tau)}{\phi_i(\tau)} = \frac{2h^2}{\pi}[\sin^{-1} \rho_i(\tau)]$$

where $\rho_i(\tau) = \phi_i(\tau)/\phi_i(0)$ and $\pm h$ are the limit levels.

Hence, to produce the required output autocorrelation function, the $\rho_i(\tau)$ to the limiter must be of the form

$$\rho_i(\tau) = \sin\left[\frac{\pi}{2h^2}\phi_0(\tau)\right]$$
$$= \sin\left(\frac{\pi}{2}\exp - \frac{|\tau|}{T}\right)$$

from eqn. (2), since $\beta = h$.

Thus $$\phi_i(\tau) = \phi_i(0) \sin\left(\frac{\pi}{2}\exp - \frac{|\tau|}{T}\right)$$

The corresponding power spectrum is obtained by expanding the series for $\sin x$ and transforming term by term, whence

$$\phi_i(\tau) = \phi_i(0)\left[\frac{\pi}{2}\exp - \frac{|\tau|}{T} - \frac{1}{3!}\left(\frac{\pi}{2}\exp - \frac{|\tau|}{T}\right)^3 + \frac{1}{5!}\left(\frac{\pi}{2}\exp - \frac{|\tau|}{T}\right)^5 - \ldots\right]$$

Using the general expression $G(\omega) = \frac{2AT}{B}\frac{1}{1 + \left(\frac{\omega T}{B}\right)^2}$ for the transform of $\phi(\tau) = A \exp - B|\tau|/T$,

$$G_i(\omega) = \phi_i(0)T$$
$$\left[\pi\frac{1}{1 + (\omega T)^2} - \frac{\pi^3}{72}\frac{1}{1 + \left(\omega\frac{T}{3}\right)^2} + \frac{\pi^5}{9600}\frac{1}{1 + \left(\omega\frac{T}{5}\right)^2} - \ldots\right] \quad . \quad . \quad . \quad . \quad (3)$$

and $\phi_i(0)$ is the input power to the non-linearity.

This may be synthesized from a source of wide-band Gaussian random signal and the outputs of appropriate linear filters of the form

$$G(\omega) = \pm\frac{G_F}{1 + \left(\omega\frac{T}{F}\right)^2}$$

These are simple single-time-constant low-pass filters with break-points at $\omega = 1/T$, $3/T$, $5/T$, etc., followed by amplifiers of gain $\sqrt{\pi}$, $-\sqrt{\pi^3/72}$, $\sqrt{\pi^5/9600}$, etc.

The required pulse train may be generated as shown in Fig. 2. The square wave is obtained by applying a Gaussian signal of the spectrum determined by eqn. (3) to a high-gain limiter. An output pulse is produced for each zero crossing of this square wave.

Although the amplitudes of successive terms tend to zero rapidly, the resultant spectrum at high frequencies is critically

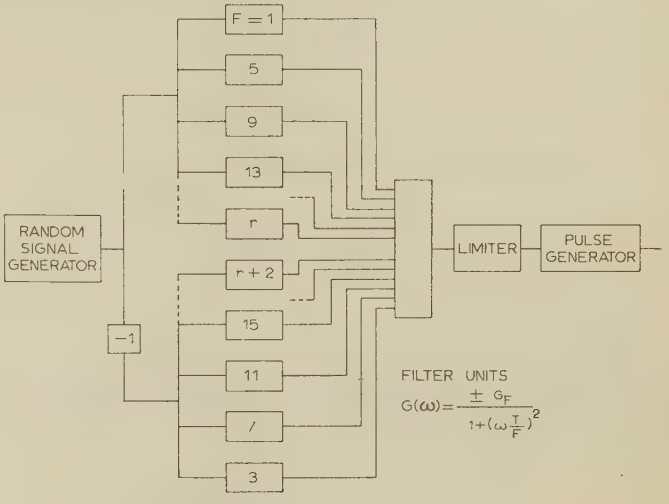


Fig. 2.—Random pulse generation by amplitude limiting of shaped spectrum.

dependent upon the ratio of cut-off frequencies of the various filters and on their various gains. Using the experimental techniques described later for determining pulse distribution, it was found that a small variation in cut-off frequency or gain of the filter produced considerable variation in the distribution of pulses and in mean rate. After considerable experiment it was decided that this method was not a practical solution to the problem.

(3) A PRACTICAL DEVELOPMENT

In view of the difficulty of obtaining a reliable Poisson pulse distribution by the system previously described, a more direct method of obtaining randomly distributed pulses has been utilized. The technique adopted is the regular sampling of a random signal, each sample producing a pulse if the instantaneous amplitude of the random signal exceeds some predetermined value.

A measure of the dependence of the amplitude of the random signal at two successive sampling instants is given by the autocorrelation function of the signal. For the amplitude of the two samples to be independent when obtained at time separation τ seconds, this autocorrelation function, $\phi(\tau)$, must approach zero.

In the practical realization of this principle the random signal is obtained by passing the output of a wide-band random-signal generator through a single-stage low-pass filter of transfer function

$$\frac{e_o}{e_i} = \frac{1}{1 + pT_i} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where $T_i = 10^{-4}$ sec.

Thus the autocorrelation function of the signal to be sampled is

$$\phi(\tau) = \phi(0) \exp - |10^4 \tau|$$

The correlation between samples separated by a time interval of 1 ms is $\phi(10^{-3}) = \phi(0)\varepsilon^{-10}$. Hence successive samples may be considered to be independent. In particular, the probability that the amplitude will exceed some value at one sampling instant is not influenced by the result of previous samples.

The mean rate is designed not to exceed 10 pulses/sec, and under these conditions the time interval between successive output pulses may be considered to be a continuous variable.

The probability of a signal having an instantaneous value in

the range v to $v + dv$ is given by the amplitude probability-density function $p(v)$. Assuming a Gaussian signal of mean square value σ^2 and with zero mean value,

$$p(v)dv = \frac{2}{\sigma\sqrt{2\pi}} \exp \frac{-v^2}{2\sigma^2} dv$$

The probability of this signal exceeding a value V is

$$\int_V^\infty p(v)dv = \frac{1}{2}\left[1 - \text{erf}\left(\frac{V}{\sigma\sqrt{2}}\right)\right]$$

where $\text{erf}\, z = 2\int_0^z \varepsilon^{-t^2} dt$.

This is the tabulated error function.[6]

Thus, if the Gaussian signal is sampled at a rate of N samples per second, the expected mean rate, ν, at which the instantaneous values of the signal exceed an amplitude V is given by

$$\nu = \frac{N}{2}\left[1 - \text{erf}\left(\frac{V}{\sigma\sqrt{2}}\right)\right]$$

This relationship, with experimental points, is shown in Fig. 3. The close agreement between theoretical and measured values

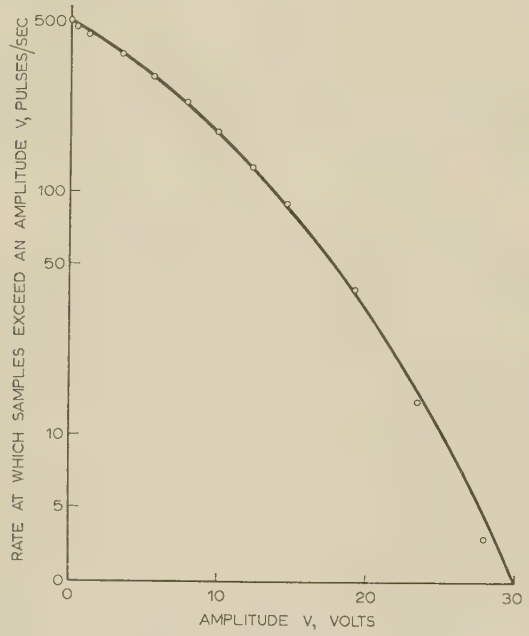


Fig. 3.—Variation of rate with amplitude.
Sampling rate = 1 000 pulses/sec.
σ^2 of random signal = $(10 \cdot 7\,\text{V})^2$.
—— Theoretical.
o o o Experimental.

indicates that the random signal used was Gaussian. It should be emphasized, however, that this is not a necessary characteristic of the random signal.

An output pulse is produced each time the random signal exceeds the defined amplitude V, and the mean output rate depends on the r.m.s. amplitude of the random signal. Since this amplitude is liable to long-term fluctuation, feedback is added to give automatic control of mean rate. This is achieved by comparing the mean value of the output pulse sequence with a preset voltage, and applying the error voltage to a high-gain integrator. The output of this integrator adjusts the test level, V, in the appropriate manner.

Fig. 4 gives the block diagram of the practical system. The

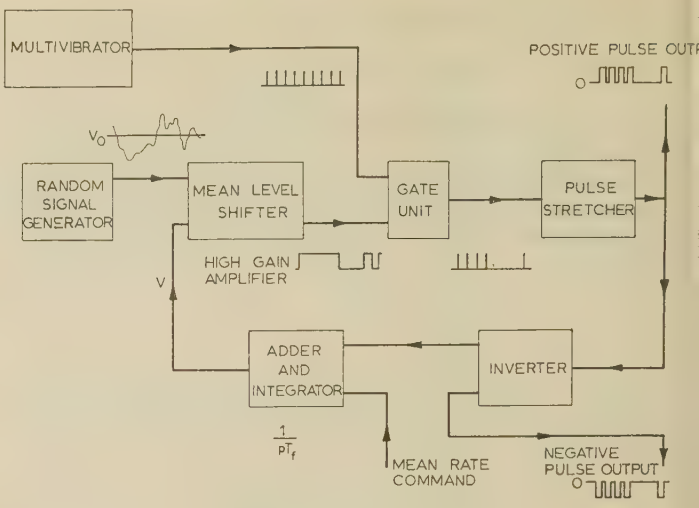


Fig. 4.—Block diagram.

primary source of the random signal is a thyratron in a magnetic field.[7] This signal has a uniform power spectrum over the audio-frequency range, and is modified by a feedback amplifier whose frequency response is determined by eqn. (4). The random signal is added to a constant voltage, V, and the polarity of the resultant signal is determined by the subsequent high-gain amplifier. The resulting waveform is sampled at a rate of 1 000 samples/sec by a short pulse (about 5 μs), producing an output pulse whenever the random signal exceeds V. For convenience, the final output pulses are of 10 ms duration and 100 V amplitude, enabling high-speed counters to be operated directly from the unit.

It is important that the time-constant of the integrator be sufficiently large, otherwise V will vary significantly under operating conditions, modifying the characteristics of the output pulse train by reducing the variance of the time intervals between successive pulses.

The effect of the integrator is to eliminate the very-low-frequency fluctuations in the output pulse train, and this effect must be negligible over a frequency range comparable with the pulse repetition frequency.

The open-loop gain of the feedback path is

$$\frac{e_o}{e_i} = K\frac{1}{j\omega T_f}$$

where T_f is the integrator time-constant and K is the gain of the remainder of the loop.

From Fig. 3, a 1 V change of V about the operating point producing 10 pulses/sec gives a change in mean rate of 0·33 pulses/sec, and since each output pulse is of 100 V ampli-

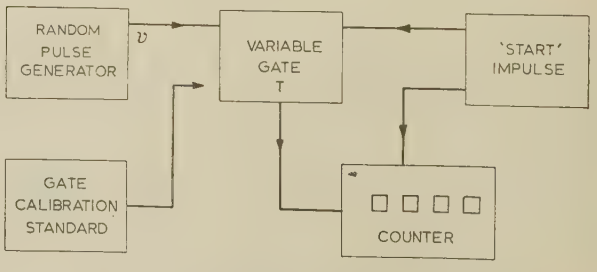


Fig. 5.—Measurement of pulse distribution.

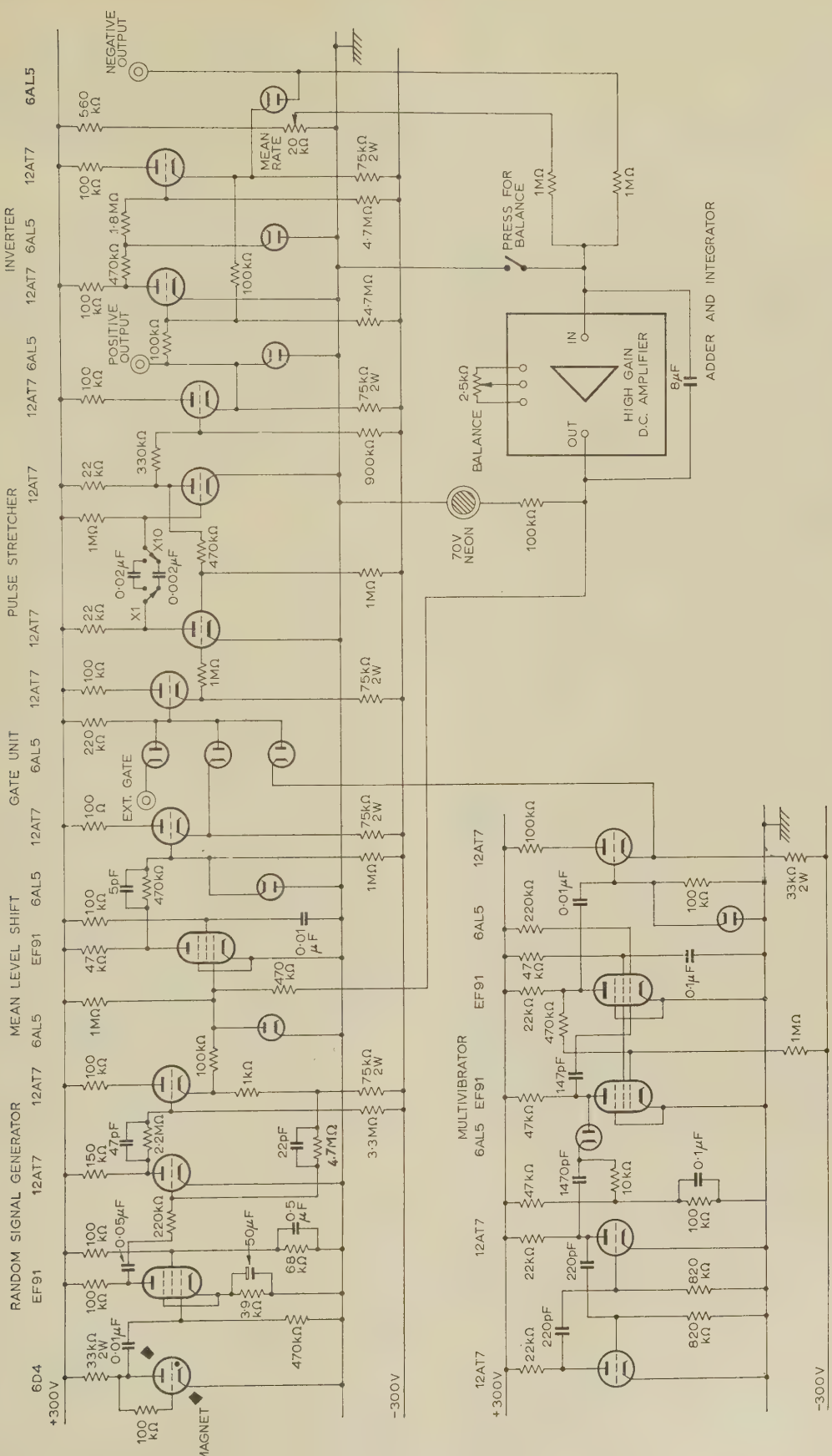


Fig. 6.—Practical circuit.

tude and of 10 ms duration, this produces a variation in average value of the output waveform of 0·33 V. Hence, at the operating point considered, $K = 0{\cdot}33$.

The closed-loop gain determines the effect on the output waveform, and this is given by

$$G(j\omega) = \frac{j\omega\frac{T_f}{K}}{1 + j\omega\frac{T_f}{K}}$$

Thus the response is -3 dB at a frequency of 0·005 c/s if $T_f = 10$ sec, producing a negligible modification of the output waveform. With this method of control the output rate is a linear function of the comparison voltage fed into the integrator, and is stable over long periods of time. The rate is readily reduced by a factor of 10 : 1 by increasing the width of the output pulses in the same ratio.

(4) MEASUREMENT TECHNIQUE

To investigate the time distribution of the pulse sequence it is required to measure the mean pulse rate and the probability distribution of a number of pulses arriving in a given time interval. An electronic counter operating up to 4 000 pulses/sec is employed in both these measurements. The mean rate is determined with high accuracy by noting the time for 10 000 pulses to occur.

The number of pulses occurring in a given time T is measured by inserting an electronic gate prior to the counter, the time during which the gate is open being measured by feeding the gate from a 1 kc/s signal generator. For each value of T, 1 000 readings are taken of the number of pulses occurring in this time, giving the experimental values of P(nT). For a particular time interval T, Fig. 1 gives the theoretical probability of occurrence of n pulses.

With the pulse generator described in Section 2 significant departure from the desired probability distribution is observed. By small adjustment of the filter characteristics, the distribution showed either (*a*) excessive rate of occurrence of zero pulses in time T, corresponding to 'bunching' of the output pules, or (*b*) too low a value of P(0, T), as may be produced by a degree of regularity in the pulse sequence.

The final pulse generator described gave results agreeing well with the expected Poisson distribution. A quantitative check of the measurement of agreement is provided by the χ^2 test.[1] This test measures the probability, Q,* that a sample of the same number of measurements from a true Poisson-event generator will produce divergencies from P(nT) greater than those observed experimentally. Values of Q greater than 0·05 are satisfactory, this figure implying that 5% of similar samples from an exact generator of Poisson distributed pulses would show greater errors than those obtained from the experimental generator.

The χ^2 test has shown that the results taken satisfy the assumption that the generator is producing pulses which obey the Poisson distribution. For measurement time intervals in the range $0{\cdot}1 < T < 10$, Q was satisfactory, and lay in the region $0{\cdot}8 > Q > 0{\cdot}1$. Hence the distribution is regarded as satisfactory.

* Q is normally denoted by the symbol P in the literature.

(5) CONCLUSIONS

Two methods of generating a random pulse sequence have been discussed, and practical considerations show that the technique of sampling a Gaussian signal produces a satisfactory result. Advantages of this method are the ease with which the mean rate is varied, and the linear relationship between mean rate and control voltage, an important consideration in some computer studies.

The sampling technique adopted in the paper is immediately applicable to the determination of the amplitude-probability distribution of signals in communication and control systems.

Fig. 6 gives details of a satisfactory practical circuit.

(6) ACKNOWLEDGMENT

The authors wish to express their thanks to Mr. J. G. Thomason, of Imperial Chemical Industries, who initiated their interest in the project, and with whom many fruitful discussions have been held.

(7) REFERENCES

(1) AITKEN, A. C.: 'Statistical Mathematics' (Oliver and Boyd, 1952).

(2) RUTHERFORD, E., and GEIGER, H.: 'The Probability Variation in Distribution of Alpha Particles', *Philosophical Magazine*, Series 6, 1910, **20**, p. 698.

(3) BENDAT, J. S.: 'Principles and Applications of Random Noise Theory' (Wiley, 1958).

(4) NEWTON, G. C., GOULD, L. A., and KAISER, J. F.: 'Analytical Design of Linear Feedback Controls' (Wiley, 1957).

(5) BARRETT, J. F., and COALES, J. F.: 'An Introduction to the Analysis of Non-Linear Control Systems with Random Inputs', *Proceedings I.E.E.*, Monograph No. 154 M, November, 1955 (**103** C, p. 190).

(6) JAHNKE, E., and EMDE, F.: 'Tables of Functions' (Dover, 1945).

(7) 'Noise Generator', *Mullard Newsletter*, July, 1955.

(8) ANDERSON, G. W., ASELTIN, J. A., MANCINI, A. R., and SARTURE, C. W.: 'A Self-Adjusting System for Optimum Dynamic Performance', *Institute of Radio Engineers National Convention Record*, 1958, Part 4, p. 182.

(9) WONHAM, W. M., and FULLER, A. T.: 'Probability Densities of the Smoothed Random Telegraph Signal', *Journal of Electronics and Control*, 1958, **4**, p. 567.

621—523 : 621.372.6

The Institution of Electrical Engineers
Paper No. 3568 M
Sept. 1961

DESCRIBING-FUNCTION EXPRESSIONS FOR SINE-TYPE FUNCTIONAL NON-LINEARITY IN FEEDBACK CONTROL SYSTEMS

By B. P. BHATTACHARYYA, B.E., C.E., M.E.

(*The paper was first received 29th August, 1960, and in revised form 11th February, 1961.*)

SUMMARY

The paper derives describing-function expressions for a sine-type functional non-linearity. In feedback control systems non-linearities of this type are encountered where synchros are used as error-sensing devices and the controlled variable undergoes large-angle variations.

The characteristic curve of the non-linearity is approximated by straight-line segments over the entire angular range from $-\pi$ to $+\pi$ radians; the slopes and break-points are obtained by minimizing the mean square error between the exact and approximate characteristics.

The describing-function expressions are useful for studying sustained oscillatory states of feedback systems comprising such elements. An example is given to illustrate this point.

LIST OF SYMBOLS

$i(t)$ = Input to the non-linearity (function of time).
$o(t)$ = Output of the non-linearity (function of time).
i_1–i_3, etc. = Values of input to the non-linearity corresponding to the break-points 1, 2 and 3, respectively, of the approximate representation.
k_1 = Initial slope of the approximate characteristics.
k_2 = Slope of the approximate characteristics after the first break-point 1, and between points 1 and 2.
i_m = Peak amplitude of the assumed sinusoidal input to the non-linearity.
ω = Angular frequency.
o_1 = Fundamental component of the output $o(t)$.
n = Order of harmonic component, used as subscript.
ϕ_1–ϕ_3, etc. = Angles in $o(t)$, the output wave resulting from a sinusoidal input to the non-linearity, corresponding to different break-points 1, 2 and 3, respectively.
G_{D1} = Symbol of the fundamental describing function.
$G_{D3} \ldots G_{Dn}$ = Third . . . nth harmonic describing function.
$\overline{\varepsilon^2}$ = Mean square value of the difference between the exact and the approximate characteristics.

(1) INTRODUCTION

During the past decade the describing-function method has been widely used to study the effects of various types of non-linearity on the performance of feedback control systems. Since its inception by Tustin[1] for studying the effects of backlash and friction, and by Kochenburger[2] for analysing relay systems, the method has been adopted and extended[3–6] to study the effects of saturation, hysteresis, deadband and similar incidental non-linearities.

It appears, however, that no attempt has been made to apply the method to studies of systems which incorporate components having certain functional relationship, trigonometric or algebraic, between the input and output quantities.

The importance of such attempts need hardly be over-emphasized. In many applications of control and computation, functional elements such as those mentioned above form part of the closed-loop system. The synchro-pair used as an error-sensing device in position-control systems constitute one typical example. The analysis of such systems is characterized by the difficulty that non-linear 'restoring force' terms appear in the differential equations which describe the system dynamics. Direct solution of the equations may become quite involved and laborious, particularly if the system is a higher-order one. Moreover, with such an approach it may not always be possible to visualize readily the changes in the nature of the system performance with changes in certain system parameters.

These inherent disadvantages of a straightforward approach with non-linear systems have resulted in the development of techniques which are relatively simple and particularly suited for studying those features of the solutions which are of specific interest in the analysis and synthesis of feedback systems. As mentioned earlier, the describing-function method is one such technique which has found wide applications for studying the sustained oscillatory behaviour of non-linear feedback control systems, and has been developed to include the effects of almost all the commonly encountered non-linearities.

The present work is intended to further the scope of application of the describing-function method to non-linear systems where the non-linearity is of a functional type. To this end, analytical expressions for the describing function of a sine-type non-linearity are derived and presented. Finally, the usefulness of the expressions is explained by means of an example.

For a sine-type non-linearity it appears that the evaluation of the exact expressions of the describing function is not feasible. Straight-line approximations are therefore made for the entire angular range from $-\pi$ to $+\pi$ radians, thereby replacing the sine characteristics by a quadrilinear (two positive slopes followed by two negative slopes in the half range) sectional scheme. With the assumed sinusoidal input the only conceivable effect of the straight-line approximation is the introduction of certain errors in the harmonic content of the output. Since in conventional describing-function analysis the effects of all higher harmonics are neglected, it may be taken that the assumption of straight-line approximation has negligible effect on the subsequent analysis. However, it is considered important to minimize the error involved by such approximation. This is done by the application of the mean-square-error criterion to obtain the best approximation. The term 'best' here implies that the average deviation over a specified range of the assumed output from the exact value is restricted to a minimum. An identical procedure was adopted by Ergin[7] for analysing certain non-linear spring-mass systems, although his subsequent analysis is quite different.

(2) DESCRIBING-FUNCTION EXPRESSIONS

(2.1) Brief Review of the Describing-Function Method

The describing-function representation of non-linearities is based upon the assumption that a sinusoidal input signal is applied to the non-linear element and the resulting output signal

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Mr. Bhattacharyya is in the Electrical Engineering Department, Bengal Engineering College, Howrah, India.

is a distorted wave having the same period as the input signal. The nature and extent of distortion is, in general, dependent on both the amplitude and the frequency of the input sinusoid. If only the fundamental component of the output wave is considered, it becomes possible to represent the non-linearity by a complex quantity whose amplitude and phase angle are given by the ratios of the amplitude and phase displacement of the fundamental of the output relative to those of the assumed input sinusoid. It is this complex quantity which describes the non-linearity in terms of the ratio of the output fundamental and the input, and which is known as the describing function for the non-linearity. In many instances, as in the present case, the fundamental of the output wave has no phase-shift relative to the input signal; the describing function is then a real number. Moreover, since the non-linearity itself involves no energy storage elements, the describing function is also independent of the frequency of input.

The describing function mentioned above is approximate, because the higher-harmonic components of the output signal are not considered. The prime justification for this approximation is the low-pass filter characteristics exhibited by most control components which usually follow the non-linearity in closed-loop control systems. Where this assumption ceases to be rigorously true, an idea of the error involved in such approximations may be obtained by Johnson's[3] detailed analysis. Although his method involves considerable computational labour, it may still be desirable to have the third- and higher-harmonic describing-function expressions. For this reason the numerical values of third-harmonic describing-function expressions are also evaluated for this non-linearity, thereby affording a basis for comparison of the relative magnitudes of the fundamental and third-harmonic components.

For investigation of stability the describing-function expressions are superimposed on an inverse Nyquist plot of the linear portion of the system (or vice versa), whence by well-known methods it becomes possible to predict the frequency and amplitude of sustained oscillations, if any, which might occur in any particular system.

(2.2) Waveshapes Resulting from Sinusoidal Input to the Non-Linear Element

The input/output relationship (static characteristics) of the non-linearity to be considered is shown in Fig. 1, together with the straight-line approximation of the non-linearity. The slopes

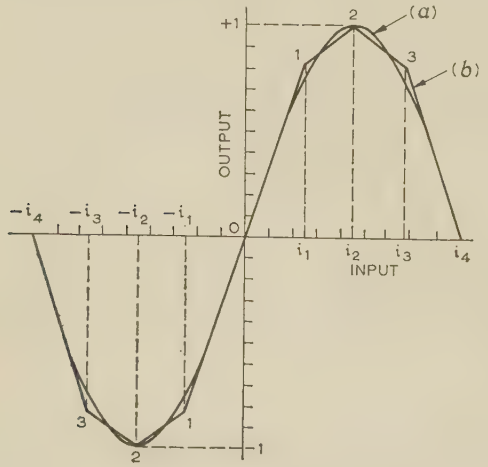


Fig. 1.—Input/output characteristics of the non-linearity.
(*a*) Exact characteristics.
(*b*) Approximate characteristics.

and break-points for best approximation, as determined on the basis of mean-square-error criterion, are derived in Section 7.

If a sinusoidal time-variation of the input $i(t)$, given by $i(t) = i_m \sin \omega t$, is assumed, the waveshapes of the resulting output, $o(t)$, will be as shown qualitatively in Fig. 2. The waveshapes depend on the range of the amplitude, i_m, of the oscillation, and the specific ranges of i_m for each case are indicated.

The mathematical expressions describing the waveshapes for various ranges of amplitude of the input sinusoid are given below. Since the waveshapes are symmetrical about the ordinates through the origin and the quarter-period time, expressions are given for a quarter-period only, the full waveshapes may be obtained by taking advantage of the symmetry.

For $0 \leqslant |i_m| \leqslant i_1$ [Fig. 2(*a*)],

$$o(t) = k_1 i_m \sin \omega t \ldots, \quad 0 < \omega t < \pi/2 \quad . \quad . \quad . \quad . \quad (1)$$

For $i_1 \leqslant |i_m| \leqslant i_2$ [Fig. 2(*b*)],

$$\left.\begin{aligned} o(t) &= k_1 i_m \sin \omega t \ldots, \quad 0 < \omega t < \phi_1 \\ &= (k_1 - k_2) i_1 + k_2 \sin \omega t \ldots, \quad \phi_1 < \omega t < \pi/2 \end{aligned}\right\} \quad . \quad (2)$$

where

$$\phi_1 = \sin^{-1} \frac{i_1}{i_m}$$

For $i_2 \leqslant |i_m| \leqslant i_3$ [Fig. 2(*c*)],

$$\left.\begin{aligned} o(t) &= k_1 i_m \sin \omega t \ldots, \quad 0 < \omega t < \phi_1 \\ &= (k_1 - k_2) i_1 + k_2 i_m \sin \omega t \ldots, \quad \phi_1 < \omega t < \phi_2 \\ &= (k_1 - k_2) i_1 + 2k_2 i_2 - k_2 i_m \sin \omega t \ldots, \\ &\qquad \phi_2 < \omega t < \pi/2 \end{aligned}\right\} \quad . \quad (3)$$

where

$$\phi_1 = \sin^{-1} \frac{i_1}{i_m}$$

and

$$\phi_2 = \sin^{-1} \frac{i_2}{i_m}$$

For $i_3 \leqslant |i_m| \leqslant i_4$ [Fig. 2(*d*)],

$$\left.\begin{aligned} o(t) &= k_1 i_m \sin \omega t \ldots, \quad 0 < \omega t < \phi_1 \\ &= (k_1 - k_2) i_1 + k_2 i_m \sin \omega t \ldots, \quad \phi_1 < \omega t < \phi_2 \\ &= (k_1 - k_2) i_1 + 2k_2 i_2 - k_2 i_m \sin \omega t \ldots, \\ &\qquad \phi_2 < \omega t < \phi_3 \\ &= (k_1 - k_2) i_1 + 2k_2 i_2 + (k_1 - k_2) i_3 \\ &\qquad - k_1 i_m \sin \omega t \ldots, \quad \phi_3 < \omega t < \pi/2 \end{aligned}\right\} \quad . \quad (4)$$

Where,

$$\phi_1 = \sin^{-1} \frac{i}{i_m}$$

$$\phi_2 = \sin^{-1} \frac{i_2}{i_m}$$

and,

$$\phi_3 = \sin^{-1} \frac{i_3}{i_m}$$

(2.3) Describing-Function Expressions

The describing function G_{D1} is given by

$$G_{D1} = \frac{\text{Fundamental amplitude of the output}}{\text{Input amplitude}}$$

Thus, for $0 \leqslant |i_m| \leqslant i_1$,

$$G_{D1} = \frac{k_1 i_m}{i_m} = k_1 \quad . \quad . \quad . \quad . \quad . \quad (5)$$

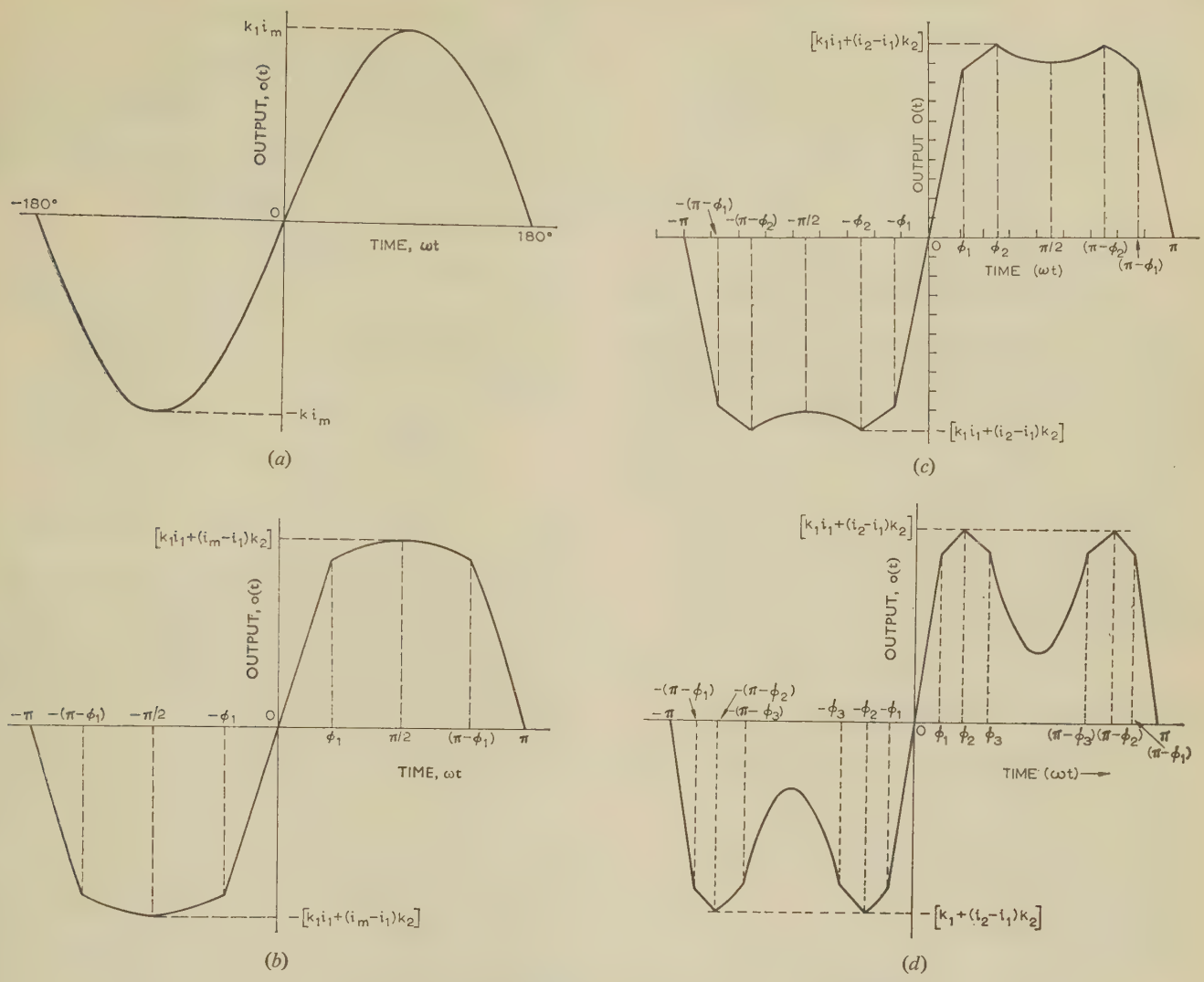


Fig. 2.—Output waves resulting from a sinusoidal input to the non-linearity.
(*a*) $0 \leqslant |i_m| \leqslant i_1$.
(*b*) $i_1 \leqslant |i_m| \leqslant i_2$.
(*c*) $i_2 \leqslant |i_m| \leqslant i_3$.
(*d*) $i_3 \leqslant |i_m| \leqslant i_4$.

For $i_1 \leqslant |i_m| \leqslant i_2$,

$$\left.\begin{aligned} o(t) &= k_1 i_m \sin \omega t, \ldots 0 < \omega t < \phi_1 \\ &= (k_1 - k_2) i_1 + k_2 i_m \sin \omega t \ldots \phi_1 < \omega t < \pi/2 \end{aligned}\right\} \quad . \quad (6)$$

The function $o(t)$ can be expanded in Fourier's series as

$$o(t) = \sum_{n=1}^{\infty} O_n \sin n\omega t$$

where O_n is the amplitude of the nth harmonic component; since $o(t)$ is an odd function, the constant term and the cosine terms of the series are absent, and only odd terms appear in the series.

Thus

$$o(t) = O_1 \sin \omega t + O_3 \sin 3\omega t + \ldots + O_n \sin n\omega t + \ldots \quad . \quad (7)$$

where n is odd.

Since the describing function is independent of frequency, we can choose $\omega = 1$ for simplicity, whence

$$o(t) = O_1 \sin t + O_3 \sin 3t + \ldots + O_n \sin nt + \ldots \quad . \quad (8)$$

where

$$O_n = \frac{2}{\pi} \int_0^{\pi} o(t) \sin ntdt$$

Taking advantage of symmetry about the ordinate through $\pi/2$, eqn. (8) is rewritten as

$$O_n = \frac{4}{\pi} \int_0^{\pi/2} o(t) \sin ntdt \quad . \quad . \quad . \quad . \quad . \quad (9)$$

Substitution of the expression for $o(t)$ from eqn. (2) into eqn. (9) yields

$$O_n = \frac{4}{\pi}\left[k_1 i_m \int_0^{\phi_1} \sin t \sin ntdt + (k_1 - k_2) i_1 \int_{\phi_1}^{\pi/2} \sin ntdt + k_2 i_m \int_{\phi_1}^{\pi/2} \sin t \sin ntdt \right] \quad . \quad (10)$$

In a similar manner, for $i_2 \leqslant |i_m| \leqslant i_3$ the nth-harmonic component of the output is given by

$$O_n = \frac{4}{\pi}\left\{ k_1 i_m \int_0^{\phi_1} \sin t \sin ntdt + \int_{\phi_1}^{\phi_2} [(k_1 - k_2) i_1 + k_2 i_m \sin t] \sin ntdt + \int_{\phi_2}^{\pi/2} [(k_1 - k_2) i_1 + 2k_2 i_2] \sin ntdt - \int_{\phi_2}^{\pi/2} k_2 i_m \sin t \sin ntdt \right\} \quad . \quad (11)$$

Identically, for $i_3 \leqslant |i_m| \leqslant \pi/2$,

$$O_n = \frac{4}{\pi}\Big\{\int_0^{\phi_1} k_1 i_m \sin t \sin ntdt + \int_{\phi_1}^{\phi_2}[(k_1 - k_2)i_1 + k_2 i_m \sin t] \sin ntdt + \int_{\phi_2}^{\phi_3}[(k_1 - k_2)i_1 + 2k_2 i_2 - k_2 i_m \sin t] . \sin ntdt + \int_{\phi_3}^{\pi/2}[(k_1 - k_2)i_1 + 2k_2 i_2 + (k_1 - k_2)i_3 - k_1 i_m \sin t] . \sin ntdt\Big\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

Evaluation of integrals given by eqns. (10)–(12) yields the amplitudes of the harmonic components. The describing-function expressions may then be obtained by dividing the respective harmonic components by the amplitude of the input sinusoid.

Thus, for $i_1 \leqslant |i_m| \leqslant i_2$,

the fundamental describing function is

$$G_{D1} = \frac{2}{\pi}\left[\frac{(k_1 - k_2)}{2}\sin 2\phi_1 + (k_1 - k_2)\phi_1 + \frac{k_2\pi}{2}\right] \quad . \quad (13)$$

the third-harmonic describing function is

$$G_{D3} = \frac{(k_1 - k_2)}{3\pi}(\sin 2\phi_1 + \tfrac{1}{2}\sin 4\phi_1) \quad . \quad . \quad (14)$$

and the nth-harmonic describing function is

$$G_{Dn} = \frac{2(k_1 - k_2)}{\pi}\left[\frac{\sin(n-1)\phi_1}{n(n-1)} + \frac{\sin(n+1)\phi_1}{n(n+1)}\right] \quad . \quad (15)$$

For $i_2 \leqslant |i_m| \leqslant i_3$

the fundamental describing function is

$$G_{D1} = \frac{2}{\pi}\left[(k_1 - k_2)\phi_1 + 2k_2\phi_2 - \frac{k_2\pi}{2} + \frac{(k_1 - k_2)}{2}\sin 2\phi_1 + k_2 \sin 2\phi_2\right] \quad . \quad (16)$$

the third-harmonic describing function is

$$G_{D3} = \frac{2}{\pi}[(k_1 - k_2)(\tfrac{1}{6}\sin 2\phi_1 + \tfrac{1}{12}\sin 4\phi_1) + k_2(\tfrac{1}{3}\sin 2\phi_2 + \tfrac{1}{6}\sin 4\phi_2)] \quad . \quad (17)$$

and the nth-harmonic describing function is

$$G_{Dn} = \frac{2}{\pi}\Big\{(k_1 - k_2)\left[\frac{\sin(n-1)\phi_1}{n(n-1)} + \frac{\sin(n+1)\phi_1}{n(n+1)}\right] + 2k_2\left[\frac{\sin(n-1)\phi_2}{n(n-1)} + \frac{\sin(n+1)\phi_2}{n(n+1)}\right]\Big\} \quad . \quad (18)$$

For $i_3 \leqslant |i_m| \leqslant i_4$

the fundamental describing function is

$$G_{D1} = \frac{2}{\pi}\left[(k_1 - k_2)\phi_1 + (k_1 - k_2)\phi_3 + 2k_2\phi_2 - \frac{k_1\pi}{2} + \frac{(k_1 - k_2)}{2}\sin 2\phi_1 + \frac{(k_1 - k_2)}{2}\sin 2\phi_3 + k_2 \sin 2\phi_2\right] \quad . \quad (19)$$

the third-harmonic describing function is

$$G_{D3} = \frac{2}{\pi}[(k_1 - k_2)(\tfrac{1}{6}\sin 2\phi_1 + \tfrac{1}{12}\sin 4\phi_1) + 2k_2(\tfrac{1}{6}\sin 2\phi_2 + \tfrac{1}{12}\sin 4\phi_2) + (k_1 - k_2)(\tfrac{1}{6}\sin 2\phi_3 + \tfrac{1}{12}\sin 4\phi_3)] \quad . \quad (20)$$

and the nth-harmonic describing function is

$$G_{Dn} = \frac{2}{\pi}\Big\{(k_1 - k_2)\left[\frac{\sin(n-1)\phi_1}{n(n-1)} + \frac{\sin(n+1)\phi_1}{n(n+1)}\right] + 2k_2\left[\frac{\sin(n-1)\phi_2}{n(n-1)} + \frac{\sin(n+1)\phi_2}{n(n+1)}\right] + (k_1 - k_2)\left[\frac{\sin(n-1)\phi_3}{n(n-1)} + \frac{\sin(n+1)\phi_3}{n(n+1)}\right]\Big\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

(2.4) Numerical Values of Describing Function for Unit Values of the Non-Linearity

The non-linear characteristic shown in Fig. 1 is given by $o = k \sin i$; assuming $k = 1$, the parameters k_1, k_2, i_1, i_2, etc., of the approximate characteristics are obtained as shown in Section 7. These are

$$\begin{aligned} k_1 &= 1 \\ k_2 &= 0{\cdot}1825 \\ i_1 &= 50° \\ i_2 &= 90° \\ i_3 &= 130° \\ i_4 &= 180° \end{aligned}$$

In the case of synchro pairs the unit of k_1 and k_2 will be the volt per radian.

When the above numerical values are substituted in the expressions previously derived the describing-function values are obtained as plotted in Fig. 3.

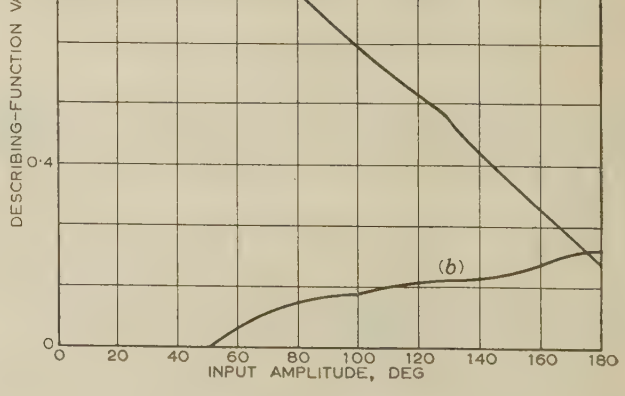


Fig. 3.—Describing-function values.
(*a*) Fundamental.
(*b*) Third harmonic.

(3) AN ILLUSTRATIVE EXAMPLE

The usefulness of the describing-function expressions derived in the previous Sections may be illustrated by considering a position-control system in which a sine-function relationship is involved by virtue of the presence of a synchro pair used as an error-detecting element. A schematic of a typical system is shown in Fig. 4. The synchro transmitter and control transformer are used to obtain an electrical signal as a sine function of the error between the reference input and load positions. Although, in the dynamic state, the output voltage of the synchro pair contains speed-induced voltages in phase quadrature with the transformer voltage, this out-of-phase voltage of the synchro pair can be eliminated by proper adjustment of the phase of the reference voltage of the phase-sensitive-detector element. The actuating error, a d.c. signal thus obtained, is power amplified and drives the amplidyne fields. The amplidyne drives the d.c. motor positioning the load.

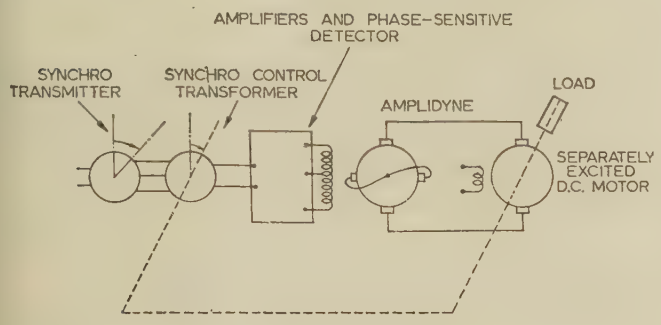


Fig. 4.—Schematic of a simple position-control servo mechanism.

Under normal conditions of operation the angular error of the position-control servo is limited to a range of few degrees, as in the case of the tracking mode of a gun-positioning system. It is then usual and valid to assume a proportional relationship between the angular error and the electrical signal of the synchro pair. But under certain conditions of operation, as when the system is synchronizing on a large step signal, the linear simplification no longer holds good. The equations describing the performance of the system being non-linear, it is no longer possible to predict the system behaviour from the simplified linear assumption. The describing-function technique then appears to be helpful in investigating the large-angle behaviour of such systems. With the describing-function expressions at hand it is possible to predict whether sustained oscillations are likely to occur, and if so, what the frequency and amplitude of such oscillations are likely to be.

To illustrate this, we assume the following numerical values of the system:

Synchro sensitivity constant	= 0·96 V/deg
Pre-amplifier gain	= 15 V/V
Power-amplifier gain	= 0·05 A/V
Amplidyne gain	= 2 200 V/A
Quadrature-axis time-constant	= 0·1 sec
D.C. motor characteristics	
Time-constant	= 1·0 sec
Gain	= 0·01 deg/sec/V

Other time-constants are considered to be negligible.

For these parameters the block diagram representation of the system is shown in Fig. 5, the normalized non-linear characteristic being placed in a separate block. The frequency-response plot for the linear portion is shown in Fig. 6, with the negative inverse values of describing function for the non-linearity superimposed. It is well known that the intersection of the frequency plot and the negative inverse plot of the describing-function values given by $G(j\omega) = -1/G_{D1}$ is a condition for increasing amplitude of oscillation. In this case, however, the point of intersection (X in Fig. 6) represents a convergent equilibrium point of sustained oscillation of constant amplitude.

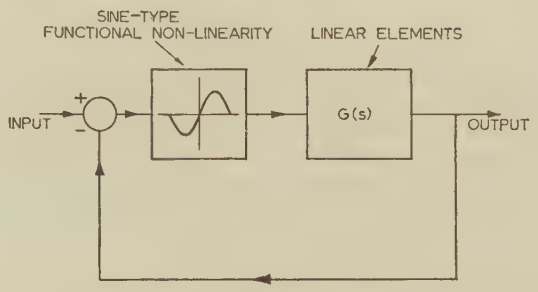


Fig. 5.—Block-diagram representation of the non-linear system.

$$G(S) = \frac{15{\cdot}84}{S(0{\cdot}1S + 1)(S + 1)}$$

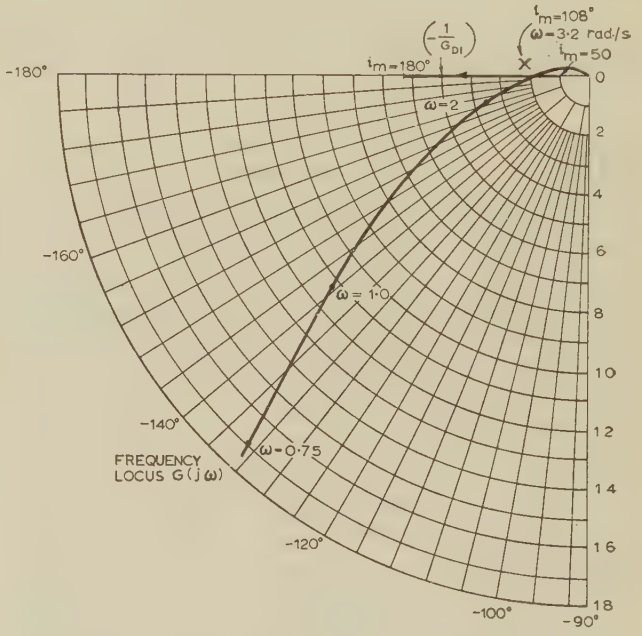


Fig. 6.—Stability analysis of the non-linear system.

The response of the system to large step signals may now be studied. If the input magnitude is less than 108°, and since $G(j\omega)$ encircles the value of $-1/G_{D1}$ (this is a consequence of the high gain of the linear portion of the system), the amplitude of oscillations tends to increase. But since the input to the non-linearity also increases at the same time, the $-1/G_{D1}$ point shifts to the left until it goes beyond the equilibrium point X. At this stage the system performance is stable, and as the oscillations tend to decrease, the $-1/G_{D1}$ point shifts to the right and the system again tends to be unstable. Under such conditions it is possible for sustained oscillations of frequency $\omega = 3{\cdot}2$ rad/sec to be maintained at an input amplitude of 108°.

(4) CONCLUSIONS

The describing-function expressions derived in Section 2 and the illustrative example given in Section 3 are indicative of the usefulness of the technique in only one of the many cases where a functional non-linearity may be involved. It is considered that, in many other applications of control and computation where functional non-linearities such as square, cubic or other trigonometrical relationship are involved, the describing function may prove to be a useful tool for studying the sustained oscillatory conditions of such systems.

(5) ACKNOWLEDGMENT

The work reported here was supported by The Council of Scientific and Industrial Research, India, under a Senior Research Fellowship Scheme awarded to the author.

The author acknowledges with thanks the co-operation received from various members of the Staff of the Electrical Engineering Department, B.E. College, Howrah, during the preparation of the paper.

(6) REFERENCES

(1) TUSTIN, A.: 'Effects of Backlash and Speed-Dependent Friction on the Stability of Closed-Cycle Control Systems', *Journal I.E.E.*, 1947, **94**, Part IIA, p. 143.

(2) KOCHENBURGER, R. J.: 'A Frequency Response Method of Analysing and Synthesizing Contactor Servomechanism', *Transactions of the American I.E.E.*, 1950, **69**, Part I, p. 270.

(3) JOHNSON, K. C.: 'Sinusoidal Analysis of Feedback Control Systems Containing Nonlinear Elements', *ibid.*, 1952, **71**, Part II, p. 169.

(4) HAAS, V. B.: 'Coulomb Friction in Feedback Control Systems', *ibid.*, 1953, **72**, Part II, p. 119.

(5) GRIEF, H. D.: 'Describing Function Method of Analysis Applied to Most Commonly Encountered Nonlinearities', *ibid.*, p. 243.

(6) SATYENDRA, K. N.: 'Describing Function Representing Effect of Inertia, Backlash and Coulomb Friction on the Stability of an Automatic Control System-1', *ibid.*, 1956, **75**, Part II, p. 243.

(7) ERGIN, E. I.: 'Transient Response of a Nonlinear System by a Bilinear Approximation Method', *Journal of Applied Mechanics* (American Society of Mechanical Engineers), 1956, **23**, Part 4, p. 635.

(7) APPENDIX

Determination of Slopes and Break-Points of the Straight-Line Approximation

For the straight-line approximations made the mean-square difference between the exact and the approximate characteristics is given by

$$\overline{\varepsilon^2} = \frac{2}{\pi}\left\{\int_0^{i_1}(k_1 i - k\sin i)^2\,di + \int_{i_1}^{\pi/2}[k_2 i + (k_1 - k_2)i_1 - k\sin i]^2 di\right\}$$

$$\text{for } 0 \leqslant |i| \leqslant \pi/2 \quad . \quad . \quad . \quad . \quad . \quad (22)$$

If k_1, the first slope of the approximate characteristics, is taken as the initial slope of the sine characteristics, the only variable on the right-hand side of eqn. (22) is i_1, the first break-point.

Hence the extreme mean square error is given by the condition

$$\frac{\partial\overline{\varepsilon^2}}{\partial i_1} = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (23)$$

Substituting eqn. (22) into condition (23) gives

$$\frac{\partial}{\partial i_1}\left\{\frac{2}{\pi}\left[\int_0^{i_1}(k_1 i - k\sin i)^2 di + \int_{i_1}^{\pi/2}[k_2 i + (k_1 - k_2)i_1 - k\sin i]^2 di)\right\}\right] = 0 \quad . \quad (24)$$

Differentiation and subsequent simplification of eqn. (24) yields

$$(k_2 - 2k_1)i_1^2 + \pi(k_1 - k_2)i_1 + \frac{\pi^2}{4}k_2 = 2k\cos i_1 \quad . \quad (25)$$

Now
$$k_2 = \frac{k - k_1 i_1}{\pi/2 - i_1} \quad . \quad . \quad . \quad . \quad . \quad . \quad (26)$$

Substitution of the non-dimensional numerical values of $k = 1$ and $k_1 = 1$ into eqn. (26), gives

$$k_2 = \frac{1 - i_1}{\pi/2 - i_1} \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

Finally, substitution of the values of k, k_1 and k_2 into eqn. (25) gives the following trigonometrical equation:

$$0{\cdot}5i_1^3 - 1{\cdot}07i_1^2 - 0{\cdot}34i_1 + 1{\cdot}235 = (1{\cdot}57 - i_1)\cos i_1 \quad . \quad (28)$$

One solution of i_1 is $0{\cdot}886$ rad $= 50°$.

This is the limit up to which the sine function may be approximated by the initial slope. The value of k_2, obtained from eqn. (27), is $0{\cdot}182\,5$.

621.317.77 : 621—52

The Institution of Electrical Engineers
Paper No. 3688 M
Sept. 1961

AN AUTOMATIC ELECTRONIC NYQUIST PLOTTER

By H. J. FRASER, Associate Member, and W. V. P. REECE.

(*The paper was first received 28th December, 1960, and in revised form 7th June, 1961.*)

SUMMARY

The instrument described is designed to plot automatically in polar co-ordinates the transfer functions of electrical networks or servo systems over the frequency range 0·2–200c/s. It consists basically of:

(*a*) A variable-frequency resistance-tuned oscillator which delivers a constant voltage to the system under test.

(*b*) A narrow-band-pass *RC* filter tracked with the oscillator to eliminate noise from the output of the system under test.

(*c*) An *RC* 90°-phase-shifting circuit tracked with the oscillator to obtain a radial measure of output voltage on an oscilloscope.

(*d*) A 12in cathode-ray display unit with pulse brightening which shows output voltage, phase and frequency in the form of a Nyquist plot as the frequency of the oscillator is varied.

Eight ganged 10-turn helical resistors are used for tracking the oscillator, filter and 90°-phase-shifting circuits.

Owing to tracking errors, the filter results in maximum inaccuracies in the plot of ±0·4dB amplitude and ±2° phase shift. The bandwidth of the filter is 20% of the operating frequency.

(1) INTRODUCTION

In the development and design of feedback servo systems and their compensating circuits, the transfer function between the input quantity and output quantity is required so that relative stability and other criteria of servo performance may be readily assessed.

A number of measuring equipments have been described in the literature by means of which point-by-point measurements of these transfer functions, given by the relations between the amplitude and phase of an output voltage compared with an input voltage, may be determined.

Early point-by-point methods of measurement described by Korn and Korn[1] and Landin[2] made use of calibrated phase-shift circuits and calibrated attenuators to obtain phase and amplitude balance between input and output voltages. The most serious limitation of these methods is that accurate measurements cannot be obtained in the presence of noise or random motion which is present in any practical servo system.

This disadvantage has led to the development of point-by-point equipments with improved signal/noise ratio based on the wattmeter principle[3] or the dynamometer phase-sensitive voltmeter described by Burns and Cooper.[4]

More recently, Paul and McFadden[5] obtain a phase-sensitive voltmeter by a time-sampling technique and they describe a method of eliminating errors due to d.c. drift and even harmonics.

The equipment described in the present paper obtains discrimination against noise by means of a new narrow-band-pass active *RC* filter which is kept in tune with the frequency of measurement by ganging together the tuning resistors of the oscillator and the filter. This filter suppresses all components of drift and noise outside its pass band, and the equipment has the advantage of automatic presentation of the transfer function by operation of a single oscillator tuning control.

A Nyquist plot is presented on the screen of a long-persistence cathode-ray tube, and the effects of changes in servo components or in the compensating networks of a servo system under test can quickly be seen. This facility permits rapid experimental adjustment of variables to obtain optimization of a servo system. The optimum shape of the Nyquist plot for a particular servo system may be obtained by immediate inspection of the changes in the plot as experimental adjustments are made to the servo components or its compensating networks.

Direct coupling is used throughout, but because the d.c. gain is only 3, drift correction is not required.

(2) DESCRIPTION OF THE EQUIPMENT

Fig. 1 shows a block diagram of the equipment. A resistance-tuned Wien-bridge oscillator (Fig. 2) is used to generate the input signals to the servo under test. This oscillator also drives a brightening-pulse generator which produces positive pulses to brighten the trace of the display cathode-ray tube (c.r.t.) at times corresponding to zero phase angle of the oscillator waveform.

The output voltage V_o from the servo under test is passed through the resistance-tuned filter (Figs. 3, 4 and 5) to eliminate noise, harmonics and random motion. This is an active *RC* filter which includes a parallel-T and a Wien-bridge circuit. It is designed to increase the amplitude of the signal frequency by 45 dB without introducing significant phase shift.

The signal output from the filter next enters a 90°-phase-shifting circuit (Fig. 6) which produces two equal voltages, one in phase with V_o and the other in quadrature with V_o.

The in-phase voltage is applied to a transistor amplifier which magnetically deflects the beam of the c.r.t. in the Y-axis, and the quadrature voltage is used similarly to obtain deflection in the X-axis. By this means the spot of the c.r.t. travels in a circular path with a radius proportional to V_o. This spot is made visible only by the brightening pulses which occur at times corresponding to each zero phase angle of the input voltage, and therefore the angular position of the brightened spot as measured from the X-axis is a direct measure of phase difference between the input and output voltages.

As the oscillator frequency is varied, the position of the brightened spot traces out the required plot.

Frequency-calibration points are obtained by blanking out the brightening pulses by switch S_2 as the oscillator is tuned through a number of calibration frequencies. These calibration points appear as small dark segments in the path of the plot. A long-persistence c.r.t. is used so that, if required, the plot can be traced directly from the screen as it is being written. The tuning resistors are hand driven through a reduction gear for flexibility as this allows a frequency-sweep speed suited to any system under test.

If no resonances exist in the system under test, a complete sweep on the high range (5–200 c/s) can be made in 20 sec without introducing errors due to the response time of the filter or the oscillator. A correspondingly longer time is required for the low-frequency range. If a large resonance appears during plotting, it is necessary to reduce the sweep speed through resonance so that an accurate plot is obtained. An analysis of this problem is given by Barber.[6]

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Mr. Fraser is with the Australian Atomic Energy Commission, and Mr. Reece is with the Sperry Gyroscope Co., Salisbury, S. Australia.

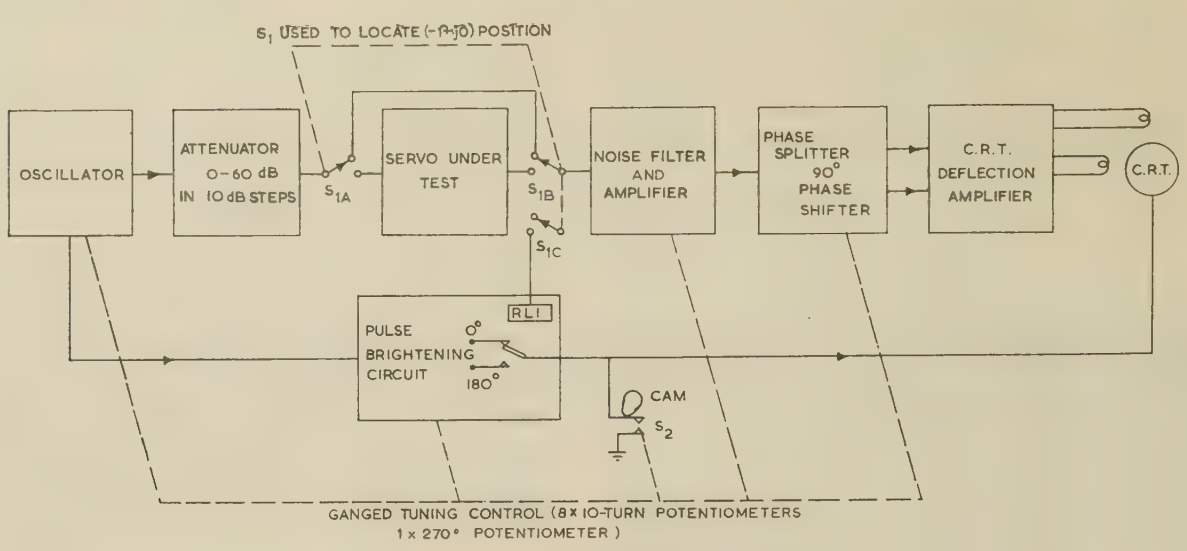


Fig. 1.—Block schematic of Nyquist plotter.

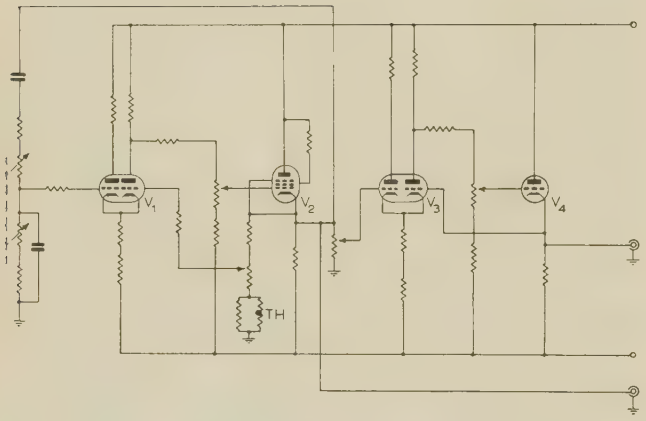


Fig. 2.—Circuit of oscillator.

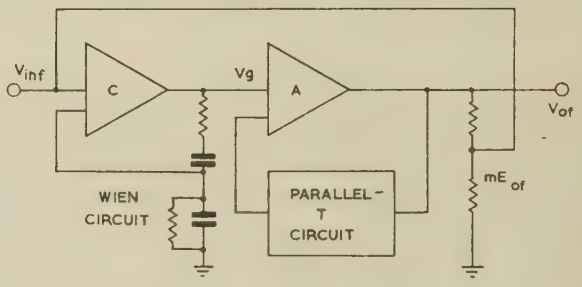


Fig. 3.—Simplified diagram of filter.

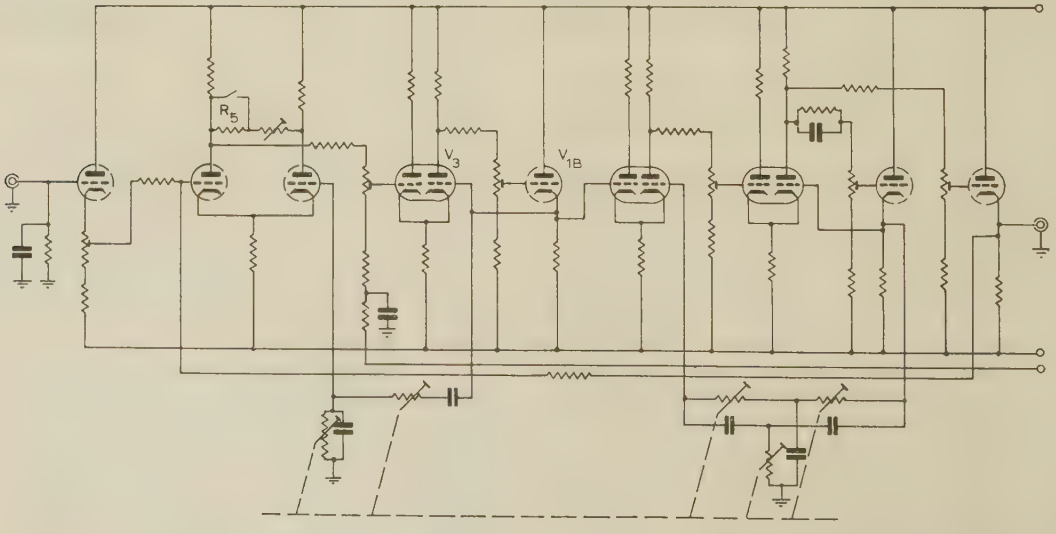


Fig. 4.—Circuit of filter.

The oscillator circuit (Fig. 2) is a conventional Wien-bridge circuit with amplitude stabilization obtained by a thermistor of suitable time-constant for the frequency range selected. It has been found practicable to cover a frequency range of 40 : 1 in a single sweep. An output impedance of 3 Ω is achieved by the augmented cathode-follower[7] V_3, V_4.

In the brightening-pulse generator which delivers positive pulses to the control grid of the c.r.t. the base of a transistor is driven by 20 V (p–p) from the oscillator (cathode of V_2, Fig. 2). This wave is squared by the transistor and valve and is then differentiated by an *RC* network, the resistive component being ganged to the oscillator tuning shaft to ensure that the output pulse width is approximately a constant fraction of a cycle and hence that the width of the c.r.t. trace is the same at all frequencies. Relay contacts are used to change over the phase of the output from 0 to 180° when required. This relay is closed by switch S_{1C} (Fig. 1) which is operated to locate the reference point $-1 + j0$.

A cam-operated switch driven by the oscillator tuning shaft blanks the brightening pulses as the oscillator is tuned through a number of calibration frequencies. If a continuous sweep is taken, these pulse-blanking intervals indicate the frequency scale of the plot.

A simplified block diagram of the filter is given in Fig. 3 and the complete circuit in Fig. 4. Referring to Fig. 3, it is seen that the first section of the filter consists of a Wien circuit connected with positive feedback around an amplifier with a gain of *C*. In normal adjustment the gain *C* is less than 3 so that this section of the circuit is not self-oscillatory. The output from this section then enters an amplifier of gain *A* which has a symmetrical parallel-T connected across it giving negative feedback. A fraction, *m*, of the output voltage V_{of} is also taken as negative feedback to the input terminal of the filter. It is shown in Section 5 that the transfer function of the filter is given by

$$\frac{V_{of}}{V_{inf}} = \frac{CA}{CAm + (1 - CY)(1 + AB)} \quad . \quad . \quad (1)$$

where *Y* is the transfer function of the Wien circuit and *B* is the transfer function of the parallel-T circuit.

Near resonance $B \simeq 0$, $Y \simeq \frac{1}{3}$ and $|(1 - CY)(1 + AB)| \simeq |(1 - CY)|$, which can be made small by a suitable choice of *C*. It can be seen from eqn. (1) that, because this latter term can be made much smaller than *CAm* for moderate values of *m*, the phase shift can also be made small with high overall gain $(1/m)$. It is possible to produce a flat-topped narrow-band-pass filter by proper selection of *C*, *A* and *m*. The method of design adopted was first to plot $(1 - CY)(1 + AB)$ for various combinations of the parameters *A* and *C*. Using these curves, an optimum value of *CAm* can be selected graphically to give a small phase shift consistent with high selectivity.

It is possible to choose *CAm* so that the vector length of $CAm + (1 + AB)(1 - CY)$ is approximately constant over the frequency range $0{\cdot}9f_0 - 1{\cdot}1f_0$, and a flat-topped response curve is obtained over this range. The values $A = 100$, $CAm = 1$ and $C = 2{\cdot}7$ have been chosen for the present equipment. Phase shift of the filter for $\pm 1\%$ deviation from f_0 is $\pm 2°$, and the amplitude change is less than 0·4 dB. Tracking errors lie within these limits.

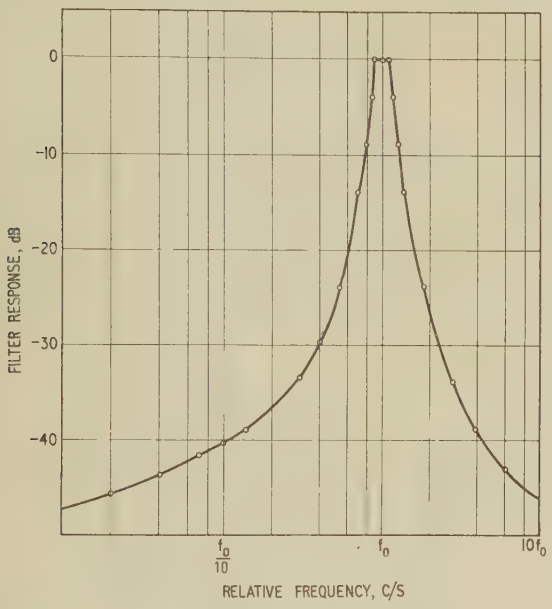


Fig. 5.—Frequency response of filter.

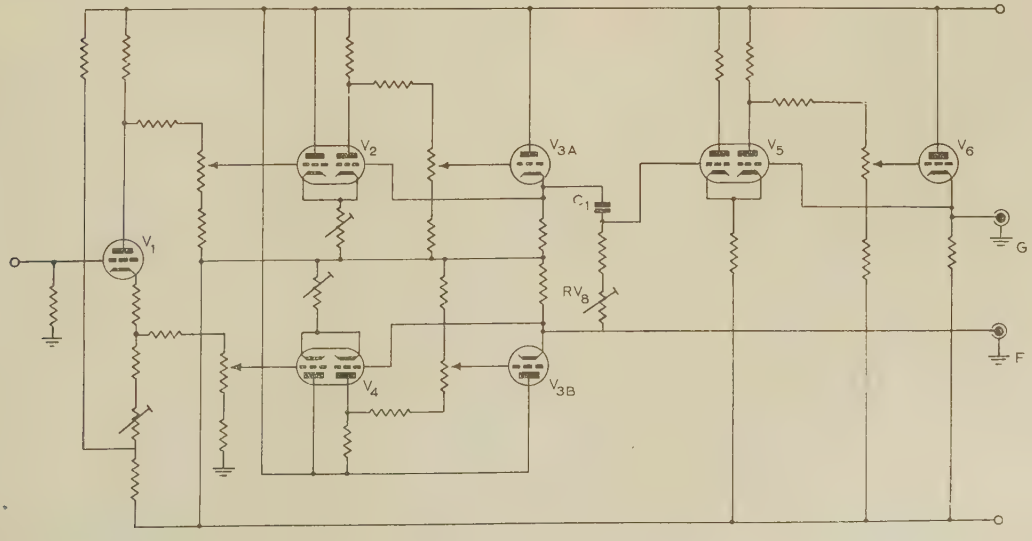


Fig. 6.—Circuit of 90° phase-shifter.

The method of setting the gain, C, of the Wien-bridge circuit is to increase C until this circuit oscillates and then reduce the gain by switching in a fixed attenuation (R_5, Fig. 4) to give $C = 2{\cdot}7$.

The filter amplitude characteristic as measured experimentally is shown in Fig. 5, from which it is seen that the desired signal is amplified by 45 dB, and at the frequencies $\frac{1}{2}f_0$ and $2f_0$ the response to noise is greater than 25 dB below the signal.

In Fig. 4 it is seen that the Wien bridge is fed from an augmented cathode-follower, V_{1b}, V_3, which has an output impedance of 3 Ω, ensuring that C remains constant over the full frequency range.

The 90°-phase-shifting circuit is shown in Fig. 6. The signal voltage from the filter enters a phase splitter, V_1, to obtain two equal voltages 180° out of phase. Each of these voltages is fed through an augmented cathode-follower with 3Ω output impedance to the phase-shifting network C_1, RV_8. RV_8 is the 10-turn helical resistor which is ganged to the oscillator tuning control so that the voltage at F is shifted 90° from the voltage at G at all frequencies. One output is taken directly from the augmented cathode-follower V_4, V_{3b} and the other from the augmented follower V_5, V_6. The low output impedance of the augmented cathode-followers ensures constant gain of the phase-shifting circuit over the full frequency range.

Plots have been taken with the equipment on simple lead and lag RC networks and also on practical servo systems. Results are obtained very rapidly, and as well as being a useful aid to the design of servo systems, it appears particularly suitable for demonstration use in teaching institutions.

(3) ACKNOWLEDGMENT

The authors thank the Sperry Gyroscope Company for permission to publish the paper.

(4) REFERENCES

(1) KORN, G. A., and KORN, T. M.: 'Modern Servomechanism Testers', *Electrical Engineering*, 1950, **69**, p. 814.
(2) LANDRIN, R.: 'Le transféromètre', *L'Onde Électrique*, 1954, **34**, p. 594.
(3) 'A Transfer Function Analyser', *Electronic Engineering*, 1955, **27**, p. 274.
(4) BURNS, D. O., and COOPER, C. W.: 'A Harmonic-Response-Testing Equipment for Linear Systems', *Proceedings I.E.E.*, Paper No. 1386 M, September, 1952 (**100**, Part II, p. 213).
(5) PAUL, R. J. A., and MCFADDEN, M. H.: 'Measurement of Phase and Amplitude at Low Frequencies', *Electronic Engineering*, 1959, **31**, p. 142.
(6) BARBER, N. F.: 'The Optimum Performance of a Wave Analyser', *ibid.*, 1949, **21**, p. 175.
(7) BILLINGTON, C.: 'A Direct-Coupled Phase-Splitter', *ibid.*, 1958, **30**, p. 480.

(5) APPENDIX

From Fig. 3,

$$V_{of} = A(V_g - BV_{of})$$

Therefore
$$V_{of}(1 + AB) = AV_g \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Similarly,
$$V_g = C(V_{inf} + YV_g - mV_{of})$$

and therefore
$$V_g = \frac{C(V_{inf} - mV_{of})}{1 - CY} \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Substituting eqn. (3) in eqn. (2) gives

$$V_{of}(1 + AB) = \frac{CA(V_{inf} - mV_{of})}{1 - CY}$$

from which

$$V_{of}\left(1 + AB + \frac{CAm}{1 - CY}\right) = \frac{CAV_{inf}}{1 - CY}$$

and
$$\frac{V_{of}}{V_{inf}} = \frac{CA}{CAm + (1 - CY)(1 + AB)}$$

621.317.733

The Institution of Electrical Engineers
Paper No. 3646 M
Sept. 1961

CHOOSING TRANSFORMER RATIO-ARM BRIDGES

By W. H. P. LESLIE, B.Sc., Member.

(*The paper was first received* 12*th October*, 1960, *in revised form* 18*th January, and in final form* 24*th April*, 1961.)

SUMMARY

Recently published papers on transformer ratio-arm bridges have given a false impression of their operating principles. The allocation of shunt and leakage inductance to each winding of a toroidal 2- or 3-winding transformer is first considered. This is then applied to the analysis of bridges for particular practical applications proposed by Lynch, Thompson, the present author and Karo. The balance conditions are compared, when experiencing shunt stray impedance, and it is shown that three of the bridges each have their most suitable applications. This does not seem likely for the last type.

LIST OF SYMBOLS

Φ_s = Stray flux.
Φ_u = Main flux.
l_3, l_4, l_5 = Leakage inductance of windings 3, 4, and 5.
L_3, L_4, L_5 = Shunt inductance of windings 3, 4, and 5.
L_3', L_4', L_5' = Total inductance of windings (e.g. $L_3' = L_3 + l_3$).
N_3, N_4, N_5 = Number of turns of windings 3, 4, and 5.
k = Constant.
M_{34} = Mutual inductance between windings 3 and 4.
r = Series resistance of winding.
z = Resistance plus leakage impedance of winding $r + j\omega l$.
Z = Total impedance of winding $z + j\omega L$.
Z_1 = Impedance to be measured.
Z_2 = Standard impedance.
Z_E = Stray shunt impedance across winding 4.
Z_F = Stray shunt impedance across winding 3.
Z_d = Shunt impedance of detector.
n = Ratio of N_4 to N_3.
r_x = Resistive component of Z_1.
L_x = Inductive component of Z_1.
r_s = Resistive component of Z_2.
L_s = Inductive component of Z_2.

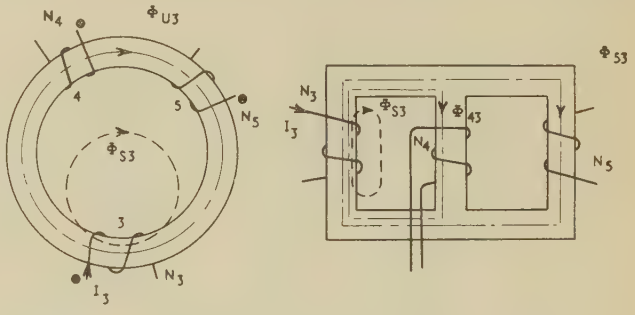
(1) INTRODUCTION

Transformer ratio-arm bridges have been gaining favour over the last few decades, the applications being the subjects of various papers.[1-9] Their use offers improvements in range, stability, freedom from errors due to earth capacitance, and a new order of accuracy. The paper discusses some of the circuits which have been proposed, determining the balance conditions involved and thus indicating the appropriate field of application of each. The discussion which has arisen concerning recent papers[8,9] indicates the need to look first at the relationship between shunt, leakage and mutual inductance in 2- and 3-winding transformers, so that analysis of the bridge circuits is based on a firm foundation.

Transformer ratio-arm bridges may have two or three windings. The effect of stray and leakage impedance is not alike in the two cases. Thompson[5] describes a typical 3-winding bridge for the comparison of two impedances, Lynch[3] a 2-winding bridge for measurement by substitution and Leslie *et al.*[7] a 2-winding bridge for comparison. Butler[9] shows that it is desirable to augment the leakage inductances of 2-winding transformers so that they are proportional to turns ratio. This result is correct despite an error in the original analysis. Karo[8] treats 2- and 3-winding transformer bridges but does not appear to have used one of them. The bridges considered by Calvert *et al.*[4] are, at first sight, different from those to be analysed in this paper because they use the transformer ratio arms to determine a current ratio rather than a voltage ratio. In general, it can be said that by interchanging voltage and detector sources a voltage-ratio bridge becomes a current-ratio bridge with identical balance conditions, provided that the same points remain earthed.

(2) INDUCTANCES OF BRIDGE TRANSFORMERS

The transformers of interest in this application have either two or three windings; the former can be treated as a special case of the latter. Three-winding transformers can be of two classes, as shown in Figs. 1(*a*) and (*b*). The type shown in Fig. 1(*a*) has



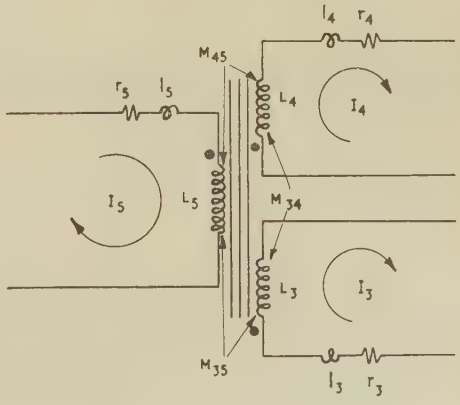


Fig. 1.—Three-winding transformer.
(*a*) Symmetrical.
(*b*) Unsymmetrical.
(*c*) Equivalent circuit.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Mr. Leslie is at the National Engineering Laboratory, Department of Scientific and Industrial Research, East Kilbride, Glasgow.

the main flux Φ_u and the leakage or stray flux Φ_s due to a winding, both symmetrically disposed relative to the other two windings. The windings may be 'lumped', as suggested diagrammatically, or symmetrically interleaved. In the type suggested by Fig. 1(*b*) the main flux due to a winding partly links one of the other windings and the rest links the other. The leakage fluxes are also asymmetrically disposed relative to the other winding.

The most suitable type of transformer for ratio-arm use is of the symmetrical type which is considered in Section 7. For our purpose each winding may be considered to have a leakage inductance, l_3, l_4 and l_5, and a shunt inductance L_3, L_4 and L_5. L_3 is not the inductance L_3', which is measured when the other windings are all open-circuit, but

$$L_3' = L_3 + l_3 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

With this definition of L_3, L_4 and L_5, it is shown in Section 7.1 that if N_3, N_4 and N_5 are the number of turns of the windings and k is a constant

$$L_3 = kN_3^2 \quad L_4 = kN_4^2 \quad L_5 = kN_5^2 \quad . \quad . \quad . \quad (2)$$

and if M_{34} is the mutual inductance between N_3 and N_4, and similarly for M_{45} and M_{53} then

$$M_{34} = kN_3N_4 \quad M_{45} = kN_4N_5 \quad M_{53} = kN_5N_3 \quad . \quad (3)$$

In the case which arises in several bridges where, say, $N_4 = N_5$

$$L_4 = kN_5^2 = L_5 = M_{45} \quad . \quad . \quad . \quad . \quad (4)$$

and this is an exact relationship.

One further problem which arises is the decision whether the voltage due to mutual inductance is to be added to, or subtracted from, the self-inductive voltage in a winding. Section 7 states a rule for this which applies to the symmetrical type of 3-winding transformer, and also defines the symbols which are used throughout the paper for winding impedance. The use of Z and z for the total winding impedance, and the winding impedance less the shunt inductance, should be noted,

Thus

$$Z = r + j\omega l + j\omega L \quad . \quad . \quad . \quad . \quad . \quad (5)$$

and

$$z = r + j\omega l \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

(3) ANALYSIS OF BRIDGES

(3.1) 'Thompson' Bridge

Thompson has made very effective use of the bridge arrangement shown in Fig, 2. which is that shown in Figs. 10 and 11 of Reference 5. There the bridge is analysed in an approximate manner and the following not-too-convenient balance equation is obtained:

$$\frac{Y_1}{Y_2} = \frac{\rho_2}{\rho_1}[1 + z_1(Y_a + Y_1) - z_2(Y_b + Y_2)] \quad . \quad . \quad (7)$$

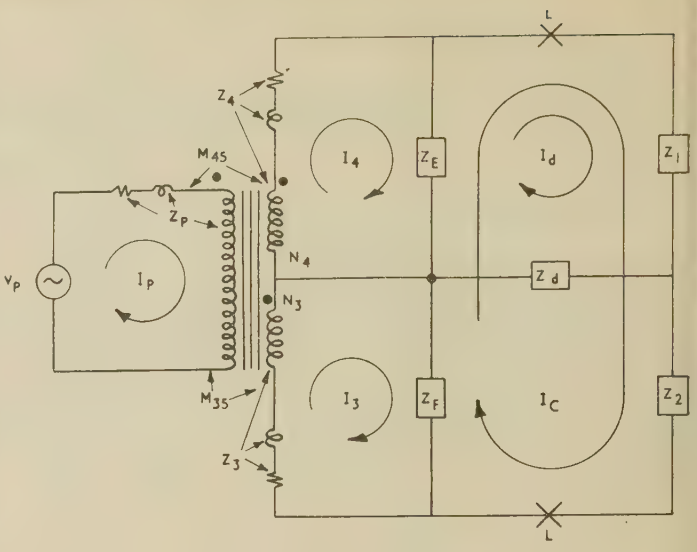


Fig. 2.—'Thompson' bridge.

Referring to Fig. 2, eqn. (7) can be written in terms of our symbols:

$$\frac{Z_2}{Z_1} = \left[1 + z_4\left(\frac{1}{Z_E} + \frac{1}{Z_1}\right) - z_3\left(\frac{1}{Z_F} + \frac{1}{Z_2}\right)\right] \quad . \quad (8)$$

This leads to rather a clumsy expression for Z_1 unless its derivation in the paper is tackled before his approximation:

$$V_1 = \frac{\rho_1 E}{1 + z_1(Y_a + Y_1)} = \rho_1 E[1 - z_1(Y_a + Y_1)] \quad . \quad (9)$$

Turning to Fig. 2 we can write the mesh equations:

	I_p	I_4	I_3	I_c	I_d	
I_p mesh	Z_p	$-j\omega M_{54}$	$-j\omega M_{53}$	0	0	$= v_p$
I_4 mesh	$-j\omega M_{54}$	$Z_4 + Z_E$	$+j\omega M_{34}$	$-Z_E$	$-Z_E$	$= 0$
I_3 mesh	$-j\omega M_{53}$	$+j\omega M_{34}$	$Z_3 + Z_F$	$-Z_F$	0	$= 0$
I_c mesh	0	$-Z_E$	$-Z_F$	$Z_1 + Z_2 + Z_E + Z_F$	$+Z_E + Z_1$	$= 0$
I_d mesh	0	$-Z_E$	0	$Z_1 + Z_E$	$Z_1 + Z_E - Z_d$	$= 0$

. . . . . (10)

At balance $I_d = 0$ and thus $\Delta_{pd} = 0$. Δ_{pd} is contained within the dotted lines.

From eqn. (3), as $N_3 = N_4$

$$M_{53} = kN_5N_3 = kN_5N_4 = M_{54} \quad . \quad . \quad . \quad (11)$$

If M_{53} is replaced by M_{54} in Δ_{pd}, which is then modified by:

(*a*) Taking first row from second row.
(*b*) Taking fourth row from third row.
(*c*) Adding third row to second row.
(*d*) Taking fourth row from second row.
(*e*) Add fourth row to first row.

giving

$$\begin{vmatrix} j\omega M_{54} & Z_4 & j\omega M_{34} & Z_1 \\ 0 & j\omega M_{34} - Z_4 & Z_3 - j\omega M_{34} & Z_2 - Z_1 \\ 0 & 0 & -Z_F & Z_2 + Z_F \\ 0 & -Z_E & 0 & Z_1 + Z_E \end{vmatrix} = 0 \quad (12)$$

this expands to give:

$$Z_F(Z_1 + Z_E)(Z_4 - j\omega M_{34}) - Z_E(Z_2 + Z_F)(Z_3 - j\omega M_{34}) - Z_E Z_F(Z_2 - Z_1) = 0 \quad . \quad . \quad . \quad (13)$$

Again from eqn. (3) because $N_3 = N_4$

$$M_{34} = kN_3N_4 = kN_3^2, \text{ or } kN_4^2$$

and from eqn. (4),

$$j\omega M_{34} = j\omega L_3 \text{ or } j\omega L_4 \quad . \quad . \quad . \quad . \quad (14)$$

Substituting in eqn. (13) we have

$$Z_F(Z_1 + Z_E)z_4 - Z_E(Z_2 + Z_F)z_3 - Z_E Z_F(Z_2 - Z_1) = 0 \quad . \quad . \quad . \quad (15)$$

This equation is the basis of Thompson's equivalent circuit for this 3-winding transformer, as shown in his Fig. 5. Eqn. (15) shows that, used in this way, the mutual inductance can be ignored provided that the total impedance of the winding is reduced by the shunt inductance term (i.e. only the series resistance and leakage reactance are used).

Eqn. (15) can be further rearranged to give

$$Z_1 = Z_2\left(1 + \frac{z_3}{Z_F} - \frac{z_4}{Z_E}\right) + \left(1 - \frac{z_4}{Z_E}\right)(z_3 - z_4) \quad . \quad (16)$$

This equation expresses Z_1 directly in terms of Z_2, showing that this arrangement is very accurate with a symmetrical transformer ($z_3 \simeq z_4$) and particularly with equal stray loads Z_E and Z_F.

(3.2) 'Lynch' Bridge

Lynch has described a bridge[3] in which use is made of the proposal in Reference 2 of applying the source directly across the largest stray impedance, so that this has no effect on balance. His bridge is shown in Fig. 3 with stray impedance Z_E and Z_F.

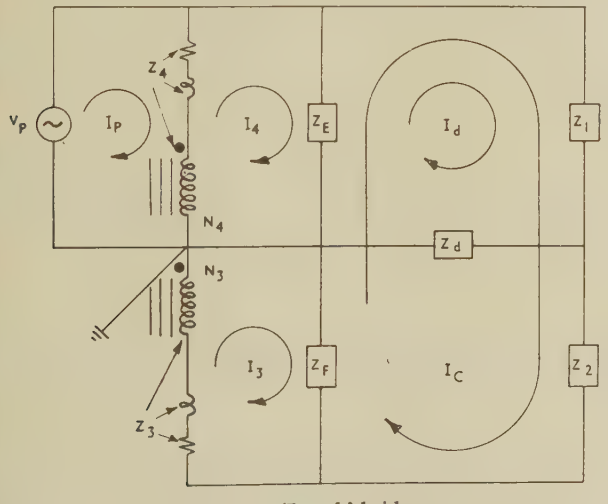


Fig. 3.—'Lynch' bridge.

Z_1 is the unknown and Z_2 is a balancing impedance. By analysis similar to that in Section 3.1 we can show that

$$Z_1 = Z_2\left(1 + \frac{z_4}{j\omega M_{34}} + \frac{z_3 + z_4}{Z_F}\right) + z_3 + z_4 \quad . \quad (17)$$

which confirms that Z_E has no effect on balance (although it loads the supply and may thus reduce sensitivity).

Lynch used this bridge by varying Z_2 for balance with an unknown connected at Z_1. He then substituted a calibrated standard at Z_1 which was adjusted for a new balance. The standard was then equal to the unknown. Used this way the bridge is excellent. It should not be used to obtain an unknown Z_1 in terms of a known Z_2 because all the errors add, the worst probably being $z_4/j\omega M_{34}$.

(3.3) 'Leslie' Bridge

The use of transformer bridges has been described in Reference 7, and that found most useful for measuring 100 Ω resistance thermometers on long leads is shown in Fig. 4. In

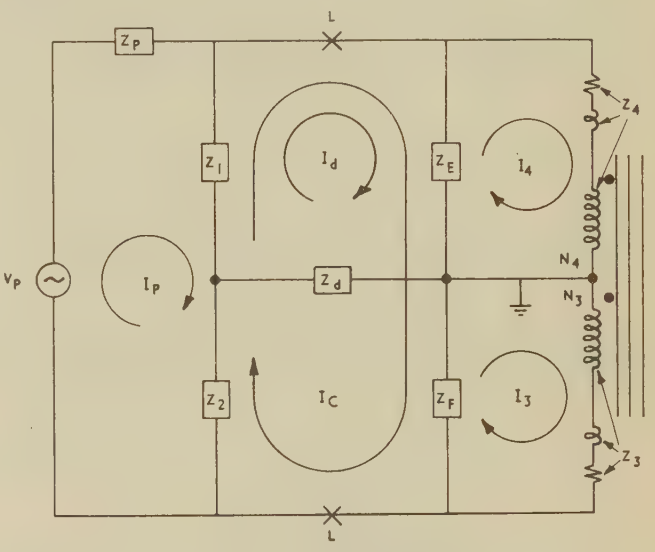


Fig. 4.—'Leslie' bridge.
$N_4 = nN_3$ at balance.

order to compare it with the similar bridges discussed in the present paper the effect of lead resistance is neglected here.

The balance condition can be shown, as before, to be

$$Z_1 = nZ_2 \frac{1 + \dfrac{z_4}{(n+1)j\omega M} + \dfrac{\frac{1}{n}z_4 + nz_3}{(n+1)Z_F}}{1 + \dfrac{nz_3}{(n+1)j\omega M} + \dfrac{\frac{1}{n}z_4 + nz_3}{\left(1 + \frac{1}{n}\right)Z_E}} \quad . \quad (18)$$

It is convenient to use approximations of the form

$$\frac{1 + a + b}{1 + c + d} = 1 + a + b - c - d$$

where a, b, c and d are all very much less than unity, to convert eqn. (18) to

$$Z_1 = nZ_2\left[1 + \frac{z_4 - nz_3}{(n+1)j\omega M} + \frac{Z_E - nZ_F}{(n+1)Z_E Z_F}\left(nz_3 + \tfrac{1}{n}z_4\right)\right]. \quad (19)$$

This form is of most use when the bridge is to be balanced by varying n as in Reference 7. To compare directly with the bridges considered previously, equal ratio arms are considered, $n = 1$, giving

$$Z_1 = Z_2\left[1 + \frac{z_4 - z_3}{2j\omega M} + \frac{Z_E - Z_F}{2Z_E Z_F}(z_3 + z_4)\right] \quad . \quad (20)$$

This shows that a good symmetrical transformer, with $z_3 \simeq z_4$

leads to high accuracy. Dependence on stray Z_E and Z_F is worst if Z_E and Z_F are very dissimilar. The third term then reduces to

$$\frac{\frac{z_3+z_4}{2}}{Z_F} \text{ when } Z_E \text{ is infinite}$$

and

$$-\frac{\frac{z_3+z_4}{2}}{Z_E} \text{ when } Z_F \text{ is infinite.}$$

This implies a potential division ratio error due to a network having the average leakage impedance as one element and the shunt stray impedance as the other (if only one shunt stray is present). When Z_E and Z_F are nearly equal, say $Z_F = Z_E(1-x)$, the

$$-\frac{\frac{z_3+z_4}{2}}{Z_E}$$

term is reduced by a factor $x/(1-x)$, where x is very small.

(3.4) 'Karo' Bridge Type 1

The 'Karo' bridge type 1 was stated to use two completely independent transformers, connected in the arrangement of Fig. 5, in which the rather odd numbering of symbols has been

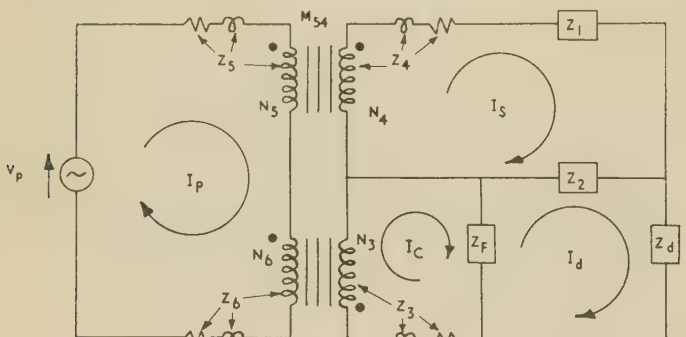


Fig. 5.—'Karo' type 1 bridge.

retained so far as convenient in order to facilitate a comparison of results. The secondary mesh currents have been given assumed directions, and the starts of windings identified in order to avoid the confusion of signs which occurred.[8] The mesh currents have been renumbered and the three elements of each winding impedance are specifically shown.

It was assumed[8] that capacitance (or any impedance) across winding 3 could be ignored. Such an impedance, Z_F, has been included in the analysis in order to check this statement.

As before, the condition for balance is

$$\frac{M_{54}}{M_{63}} = \frac{Z_1+Z_2+Z_4}{Z_2} \times \frac{Z_F}{Z_3+Z_F} \quad . \quad . \quad (21)$$

This balance equation can be considered in two stages. The multiplying factor $Z_F/(Z_3+Z_F)$ shows that Z_F has an effect on balance unless the factor takes the value of unity which occurs when Z_F is infinite. It is necessary to consider a practical transformer, which has been avoided in Reference 8; a practical audio-frequency transformer might have a shunt inductance $L_3 = 10$ H, which at a frequency of $5000/\pi$ c/s (for convenience) gives a value

$$Z_3 = j2\pi fL = j10^5 \text{ ohm}$$

A practical value of Z_F might be, for 100 pF,

$$Z_F = -\frac{1}{j2\pi fc} = -j10^6 \text{ ohm}$$

For these assumed values

$$\frac{Z_F}{Z_3+Z_F} = \frac{10}{9}.$$

Karo has stated[8] that Z_F can be neglected, and later that the effect of Z_F (his C_2+C_3) can be treated as an alteration in the value of Z_3. The equations which he claimed to be correct have no Z_3 in them, so that an alteration in Z_3 (even if it could be calculated) would have no effect due to his incorrect analysis.

Apart from this necessary correction factor we obtain from eqn. (21)

$$\frac{M_1}{M} = \frac{Z_1+Z_2+Z_4}{Z_2} \quad . \quad . \quad . \quad . \quad . \quad (22)$$

in which we have written $M_{54} = M_1$ and $M_{63} = M$ to agree with the notation in Reference 8. This is eqn. (2) of Reference 8 but we have shown that

$$Z_4 = r_4 + j\omega l_4 + j\omega L_4$$

and for any normal transformer design $j\omega L_4$, the shunt reactance, will be very much larger than the leakage reactance $j\omega l_4$ or the series resistance r_4.

Having shown that Z_4 in eqn. (22) is indeed mainly a shunt inductance, the usefulness of the type 1 bridge can be assessed when measuring an unknown Z_1 in terms of a known Z_2 and a measured (or calculated) M_1 and M.

Eqn. (22) can be written

$$Z_1 = \frac{M_1-M}{M}Z_2 - Z_4 \quad . \quad . \quad . \quad . \quad . \quad (23)$$

Resubstituting $M_1 = M_{54} = k_1N_5N_4$ from eqn. (3) and

$$M = M_{63} = kN_6N_3$$

and for identical primaries and secondaries $N_5 = N_6$ and $k_1 = k_2$; then

$$Z_1 = \frac{N_4-N_3}{N_3}Z_2 - Z_4 \quad . \quad . \quad . \quad (24)$$

Consider an unknown $Z_1 = r_x + j\omega L_x$ measured in terms of a standard $Z_2 = r_s + j\omega L_s$. The imaginary part of eqn. (24) gives

$$L_x = \frac{N_4-N_3}{N_3}L_s - (l_4+L_4) \quad . \quad . \quad . \quad (25)$$

As L_4, the transformer shunt inductance, is a large iron-cored inductance this arrangement is of little practical value.

(3.5) 'Karo' Bridge, Type 2

Fig. 6 shows the bridge of Fig. 2 of Reference 8, where it is erroneously assumed that the analysis of the type 1 bridge could be directly applied to this type 2 bridge on the ground that the bridges obeyed identical conditions. The stray impedances Z_E and Z_F have been added so that their effect can be assessed.

The condition for balance is found to be

$$Z_1 \simeq Z_2\left(1+\frac{2z_3}{Z_F}-\frac{2z_4}{Z_E}\right) - \left(1-\frac{z_4}{Z_E}\right)z_4 \quad . \quad (26)$$

This relationship between the unknown Z_1 and the standard Z_2 shows that only the resistance and leakage inductance are

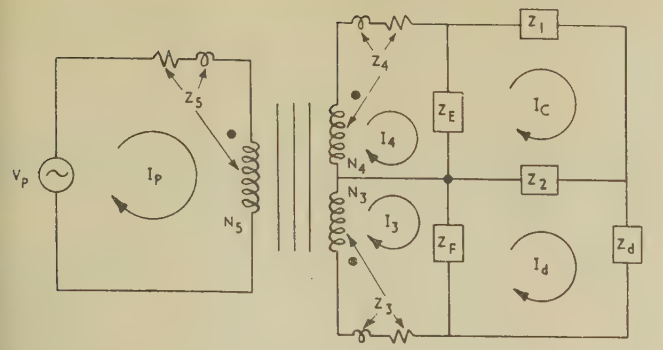


Fig. 6.—'Karo' type 2 bridge.

responsible for error, there being no shunt-inductance term. To simplify the issue, consider a bridge with Z_E and Z_F infinite, no shunt impedances, and only looking at quadrature components; eqn. (26) gives

$$L_x = L_s - l_4 \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

In the absence of information in Reference 8 we may assume that Karo used transformers with a leakage inductance l_4 of approximately 1 mH. Consider the measurement in Section 3.1.1 of Reference 8, where an unknown inductance L_x was found to be 1·000 1 mH.

Therefore

$$L_s = L_x + l_4 = 2{\cdot}000\,1\,\text{mH} \quad . \quad . \quad . \quad . \quad (28)$$

and the leakage inductance l_4 is as important as the standard inductance L_s, both requiring to be known to an accuracy of within 0·1 μH.

(4) COMPARISON OF BRIDGES

(4.1) As Considered Above

The balance conditions for the bridges considered can be contrasted with interest:

'Thompson' bridge

$$Z_1 = Z_2\left(1 + \frac{z_3}{Z_F} - \frac{z_4}{Z_E}\right) \quad \left(1 - \frac{z_4}{Z_E}\right)(z_3 - z_4) \quad . \quad (29)$$

'Lynch' bridge

$$Z_1 = Z_2\left(1 + \frac{z_4}{j\omega M_{34}} + \frac{z_3 + z_4}{Z_F}\right) + z_3 + z_4 \quad . \quad (30)$$

'Leslie' bridge

$$Z_1 = Z_2\left[1 + \frac{z_4 - z_3}{j\omega M} + \frac{(Z_E - Z_F)}{Z_E Z_F}\frac{(z_3 + z_4)}{2}\right] \quad . \quad (31)$$

*'Karo' type 1 bridge**

$$Z_1 = Z_2 - (z_4 + j\omega L_4) \quad . \quad . \quad . \quad . \quad (32)$$

'Karo' type 2 bridge

$$Z_1 = Z_2\left(1 + \frac{2z_3}{Z_F} - \frac{2z_4}{Z_E}\right) - z_4\left(1 - \frac{z_4}{Z_E}\right) \quad . \quad (33)$$

(i) For practical transformers z_3 and z_4 will be less than one-thousandth of terms like $j\omega L_4$ and $j\omega M$ and also possibly of Z_E and Z_F.

(ii) The 'Karo' type 1 bridge has a $j\omega L_4$ term which will often be larger than Z_1 and Z_2 and seldom able to be ignored. This bridge has little practical value and is quite different in principle from any of the others.

(iii) Of the remaining bridges the 'Lynch' bridge is the only one in which all the error terms are additive, making it undesirable for comparing Z_1 with Z_2. It is, however, the only bridge in which Z_E does not affect balance and was selected by Lynch for this reason. He used a substitution method in which Z_1 was successively the unknown and the standard, and Z_E was the only stray to change during this process. For such use it is the best arrangement.

(iv) The 'Karo' type 2 bridge is unfortunate in that even when Z_E and Z_F are infinite (i.e. not there) the balance condition is

$$Z_1 = Z_2 - z_4 \quad . \quad . \quad . \quad . \quad . \quad (34)$$

so that Z_1 and Z_2 must be very large compared with z_4 for accuracy.

Z_E and Z_F lead to other errors which are far from cancelling. There appears little use for this arrangement in practice.

(v) Thompson states on p. 250 of Reference 5 that where the supply is connected across the ratio arms [as for eqns. (30) and (31)] 'such ratios are not as stable as those obtained by using a separate excitation winding . . .' (as in eqn. (29)]. Some justification for this can be obtained from the fact that there is a correction term $z_3 + z_4$ in eqn. (30) and

$$Z_2\frac{Z_E - Z_F}{Z_E Z_F}\left(\frac{z_3 + z_4}{2}\right)$$

in eqn. (31), whereas in eqn. (29) the corresponding term is

$$\left(1 - \frac{z_4}{Z_E}\right)(z_3 - z_4)$$

Thus, when the transformer is designed symmetrically, $z_3 \simeq z_4$, and this error term for the 'Thompson' bridge is small.

There is, however, another error term which in eqn. (29) is

$$\frac{z_3}{Z_F} - \frac{z_4}{Z_E}$$

and, assuming $z_3 = z_4$, this term only vanishes under the special circumstance when the stray impedance Z_E and Z_F are equal. This condition is not to be relied on, but would also eliminate the error term discussed above for eqn. (31). The first error term in eqn. (31) vanishes when $z_3 = z_4$.

The conclusion is that there is little difference between the bridges corresponding to eqns. (29) and (31), both of which offer a stable and true ratio when $z_3 \simeq z_4$ and $Z_E \simeq Z_F$. In the likely practical condition when $z_3 \simeq z_4$ but Z_E is not equal to Z_F, let us assume $Z_E \gg Z_F$, and we have, from eqn. (29),

$$Z_1 \simeq Z_2\left(1 + \frac{z_3}{Z_F}\right) \quad . \quad . \quad . \quad . \quad . \quad (35)$$

and from eqn. (31) $\quad Z_1 \simeq Z_2\left(1 + \frac{z_3}{Z_F}\right) \quad . \quad . \quad . \quad . \quad (36)$

resulting in equal errors in both arrangements.

(4.2) Considering Series Lead Impedance

The entire discussion assumes that the series resistances and inductances of the leads in the bridge are negligible. The present author was interested[7] in the case when Z_1 and Z_2 were of the order of 100 Ω and at a distance such that the impedance of reasonable leads was about 1 Ω. It was for this reason that the bridge of Fig. 4 was chosen, since 1 Ω inserted at L produced attenuation errors of the order of 1 part in 10^6 in feeding the shunt inductance of N_3 and N_4 and 1 part in 10^5 in feeding

* As this bridge is unlikely to be of practical interest the effect of stray impedance is not shown; eqn. (32) assumes $Z_E = Z_F = \infty$.

Z_E and Z_F. If the bridge of Fig. 2 had been used, 1 Ω inserted at L would have altered Z_1 and Z_2 to $(Z_1 + 1)$ and $(Z_2 + 1)$, where Z_1 and Z_2 are about 100 Ω, and error would rapidly be incurred as Z_1 and Z_2 departed from exact equality. In these special circumstances (resistance thermometry or resistance strain gauging) the arrangement of Fig. 4 is best.

(5) CONCLUSIONS

The arrangements suggested by Karo[8] appear to offer no practical advantage, and the printed discussion indicates a difference of opinion on his analysis.

Of the six bridges considered, those of Figs. 2 and 4 are the most suitable for comparing similar impedances which have appreciable capacitance to earth from each terminal. That of Fig. 4 is most suitable where low impedances have to be compared at such a distance that lead impedance is not negligible. For measurement by substitution the bridge of Fig. 3 should be used.

(6) REFERENCES

(1) KIRK, H. L.: 'Radio Section: Chairman's Address', *Journal I.E.E.*, 1945, **92**, Part III, p. 2.

(2) CLARK, H. A. M., and VANDERLYN, P. B.: 'Double-Ratio A.C. Bridges with Inductively-Coupled Ratio Arms', *Proceedings I.E.E.*, Paper No. 742 M, January, 1949 (**96**, Part III, p. 189).

(3) LYNCH, A. C.: 'A Bridge Network for the Precise Measurement of Direct Capacitance', *ibid.*, Paper No. 2209 M, October, 1956 (**104** B, p. 363).

(4) CALVERT, R., CORNELIUS, J. A., GRIFFITHS, V. S., and STOCK, D. I.: 'The Determination of the Electrical Conductivities of some Concentrated Electrolyte Solutions using a Transformer Bridge', *Journal of Physical Chemistry*, 1958, **62**, p. 47.

(5) THOMPSON, A. M.: 'The Precise Measurement of Small Capacitances', *Transactions of the Institue of Radio Engineers, Instruments*, 1958, p. 245.

(6) MCGREGOR, M. C., *et al.*: 'New Apparatus at the National Bureau of Standards for Absolute Capacitance Measurement', *ibid.*, p. 253.

(7) LESLIE, W. H. P., HUNTER, J. J., and ROBB, D.: 'Precision Temperature Measurement Outside the Laboratory', *Research*, 1960, **13**, p. 250.

(8) KARO, D.: 'A Novel High Accuracy Circuit for the Measurement of Impedance in the A.F., R.F. and V.H.F. Ranges', *Proceedings I.E.E.*, Paper No. 2709 R, November, 1958 (**105** B, p. 505). Discussions: **106** B, p. 435 and **107** B, p. 225.

(9) BUTLER, F.: 'A.C. Bridges with Inductive Ratio Arms', *Electronic Technology*, 1960, **37**, p. 303, and Discussion, 1960, **37**, p. 434.

(7) APPENDIX

Three-Winding Transformer Inductances

We are concerned with transformers having two windings and three windings. The former can be considered as a special case of the latter. The 3-winding transformer is shown diagrammatically in Figs. 1(*a*) and (*b*) in order to emphasize that all 3-winding transformers do not conform to the same general performance. The unsymmetrical transformer of Fig. 1(*b*) has a more complicated relationship of mutual and self-inductance than the symmetrical transformer of Fig. 1(*a*) because in Fig. 1(*b*) the flux due to current in winding 3, for example, couples windings 4 and 5 in quite distinct proportions. In Fig. 1(*a*) the flux in windings 4 and 5 due to current 3 is identical.

Fortunately we are concerned with symmetrical transformers with toroidal cores as in Fig. 1(*a*), where windings 3, 4 and 5 may be wound simultaneously by three wires lying side by side on the core, or lumped as suggested diagrammatically, or a combination of both. Fig. 1(*a*) shows two distinct components of flux due to current I_3, namely, Φ_{s3}, the stray flux [which couples winding 3 only and is responsible for the leakage inductance shown as l_3 in Fig. 1(*c*)] and Φ_{u3} the useful flux [which also couples windings 4 and 5 and is responsible for the shunt inductance of the winding 3 shown as L_3 in Fig. 1(*c*)].

It should be realized that the inductance of winding 3, which would be measured with windings 4 and 5 open-circuited consists of $l_3 + L_3 = L_3'$, say. L_3' is usually referred to as the shunt inductance, but will not be so used here since the L_3 inductance is a more suitable measure for this analysis. Similar inductances exist for the other windings, and the total flux in the core can have six components, Φ_{s3}, Φ_{u3}, Φ_{s4}, Φ_{u4}, Φ_{s5} and Φ_{u5}, where the second subscript identifies the winding responsible for the flux.

Mutual inductance exists between each pair of windings and will be written M_{43}, where the second subscript 3 identifies the winding responsible for the flux inducing an e.m.f. in the winding 4 identified by the first subscript. A typical relation for the voltage to be found in winding 3 is

$$v_3 = I_3(r_3 + j\omega l_3 + j\omega L_3) + I_4 j\omega M_{34} + I_5 j\omega M_{35} \quad . \quad (37)$$

The question arises with each mutual inductance as to whether $a+$ or $-$ is to be used with the $I_4 j\omega M_{34}$, for example. A simple rule may be stated, once corresponding ends of windings have been identified and marked with a dot:

If I_3 and I_4 both enter 'dot' ends or both enter unmarked ends of their windings 3 and 4, the same sign is used for the self- and mutual-inductance terms, otherwise opposite signs are used.

Eqn. (37) can be shortened by writing

$$z_3 = r_3 + j\omega l_3 \text{ and similarly for } z_4, z_5, \text{ etc.} \quad . \quad . \quad . \quad . \quad (38)$$

or

$$Z_3 = r_3 + j\omega l_3 + j\omega L_3 \text{ and similarly for } Z_4, Z_5, \text{ etc.} \quad . \quad (39)$$

giving

$$v_3 = I_3 Z_3 + I_4 j\omega M_{34} + I_5 j\omega M_{35} \quad . \quad . \quad . \quad (40)$$

and similarly for v_4 and v_5.

A similar set of equations can be built up by considering the total flux in each winding. For example, for winding 3 the total flux is

$$\Phi_3 = \Phi_{s3} + \Phi_{u3} + \Phi_{u4} + \Phi_{u5}$$

and this flux, which is alternating, induces an e.m.f.

$$v_3 = N_3 \frac{d\Phi_3}{dt} = N_3\left(\frac{d\Phi_{s3}}{dt} + \frac{d\Phi_{u3}}{dt} + \frac{d\Phi_{u4}}{dt} + \frac{d\Phi_{u5}}{dt}\right) \quad . \quad (41)$$

and similarly for v_4 and v_5.

Comparing eqns. (37) and (41) term by term, remembering that $I_3 r_3$ does not appear in eqn. (41) and that $I_3 j\omega L_3$ may be interpreted as $L_3 \frac{dI_3}{dt}$ we can derive Table 1.

Table 1

$$l_3 = N_3 \frac{d\Phi_{s3}}{dI_3} \quad L_3 = N_3 \frac{d\Phi_{u3}}{dI_3} \quad M_{34} = N_3 \frac{d\Phi_{u4}}{dI_4} \quad M_{35} = N_3 \frac{d\Phi_{u5}}{dI_5}$$

and also

$$l_4 = N_4 \frac{d\Phi_{s4}}{dI_4} \quad L_4 = N_4 \frac{d\Phi_{u4}}{dI_4} \quad M_{43} = N_4 \frac{d\Phi_{u3}}{dI_3} \quad M_{45} = N_4 \frac{d\Phi_{u5}}{dI_5}$$

$$l_5 = N_5 \frac{d\Phi_{s5}}{dI_5} \quad L_5 = N_5 \frac{d\Phi_{u5}}{dI_5} \quad M_{53} = N_5 \frac{d\Phi_{u3}}{dI_3} \quad M_{54} = N_5 \frac{d\Phi_4}{dI_4}$$

Consider a relation from Table 1, such as

$$M_{34} = N_3 \frac{d\Phi_{u4}}{dI_4} \quad . \quad . \quad . \quad . \quad . \quad (42)$$

The flux change for a given current change, in a linear circuit, can be written $d\Phi_{u4}/dI_4 = k_4 N_4$, where k_4 depends on permeability, coil shape, etc., and substituting in eqn. (42),

$$M_{34} = k_4 N_3 N_4 \quad . \quad . \quad . \quad . \quad . \quad (43)$$

Similarly, $$M_{43} = N_4 \frac{d\Phi_{u3}}{dI_3} = k_3 N_3 N_4 \quad . \quad . \quad . \quad . \quad (44)$$

but reciprocity requires that $M_{34} = M_{43}$ and M_{34} would normally be written.

From eqns. (43) and (44) it is seen that $k_3 = k_4 = k$, say, and by comparing M_{35} with M_{53} we find $k_5 = k_3 = k$, so that we can write

$$M_{34} = kN_3N_4 \quad M_{45} = kN_4N_5 \quad M_{35} = kN_3N_5 \quad . \quad (45)$$

Similarly, by substituting $d\Phi_{u3}/dI_3 = kN_3$ in the value for L_3 in the Table

$$L_3 = kN_3^2 \quad \text{and} \quad L_4 = kN_4^2 \quad L_5 = kN_5^2 \quad . \quad (46)$$

It is worth reiterating that all these relations apply only to a symmetrical 3-winding transformer or to a 2-winding transformer with I_5 and $N_5 = 0$, say.

In the special case where $N_3 = N_4 = N$, say,

$$M_{34} = kN^2 \quad L_3 = kN^2 \quad L_4 = kN^2 \quad . \quad (47)$$

Therefore $$L_3 = L_4 = M_{34} \text{ exactly} \quad . \quad . \quad . \quad . \quad (48)$$

AN ANALYSIS OF THE TRAVELLING-WAVE CAVITY

By N. Karayianis and C. A. Morrison.

(*Communication first received 17th March, and in revised form 25th May*, 1961.)

An experimental inquiry into the response of a travelling-wave cavity by the authors showed that the actual response of a typical cavity differs from that predicted by the simple theory. A theoretical analysis shows that an internal mismatch can explain the observed response. It is further shown that an isolator in the cavity greatly reduces the effect of any mismatch so that the response of the cavity will deviate insignificantly (in the case considered) from the originally expected response.

List of Principal Symbols

S_0 = Scattering matrix for input directional coupler.
s = Transmission coefficient for input coupler.
a = Coupling coefficient for input directional coupler.
ϕ = Phase shift of wave after traversing cavity.
β = Damping coefficient of cavity.
η = Characteristic loss of the cavity.
δ = Phase shift from cavity resonance.
K = Cavity constant.
P_0 = Ratio of off-resonance to on-resonance power.
S_2 = Scattering matrix for mismatch in cavity.
τ = Transmission coefficient of mismatch.
ρ = Reflection coefficient of mismatch.
I = Scattering matrix for isolator in cavity.
α = Reverse transmission coefficient for isolator.

A number of articles in the literature have been concerned with the characteristics of travelling-wave cavities,[1,2] and with the applications of these cavities to microwave measurements.[3,4] These articles deal chiefly either with the forward-wave properties of the travelling-wave cavity, or with the response of the cavity when the coupling to it is non-directional.

In addition, the effect of an internal mismatch on the input impedance and the power amplification of a travelling-wave cavity has been investigated.[5]

The resulting equations describing the cavity with the mismatch are so involved that it would be tedious to use them in analysing the results of any measurement. The effect of an ideal isolator placed in the cavity with mismatch was then considered and is shown to simplify the equations considerably.

The authors are at the Diamond Ordnance Fuze Laboratories, Washington, D.C., U.S.A.

The Theory

The Simple Cavity.—A travelling-wave cavity in its simplest form consists of a closed loop of waveguide containing as one of its elements a directional coupler. Microwave energy is coupled into the loop through the directional coupler. To analyse the response of what will be called the 'simple cavity' we assume a lossless directional coupler with perfect directivity. The symbolic and schematic representation of such a coupler is shown in Fig. 1, where S_0 is the scattering matrix such that

$$S_0 \equiv \begin{pmatrix} 0 & a & 0 & js \\ a & 0 & js & 0 \\ 0 & js & 0 & a \\ js & 0 & a & 0 \end{pmatrix}$$

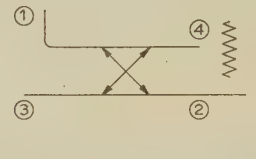


Fig. 1.—Scattering matrix for the input directional coupler.

$E_{out} = S_0 E_{in}$, the E's are column matrices, and s and a are the transmission and coupling coefficients respectively such that $a^2 + s^2 = 1$. The attenuation and phase shift of the microwave amplitude in the cavity is represented by the factor $\exp\left(j\frac{\phi}{l}x - \beta x\right)$, where x is the length travelled by the wave, l is the total length of the cavity, ϕ is the total phase shift around the cavity and β is the damping coefficient. The characteristic loss of the cavity is defined as $\eta = \varepsilon^{-\beta l}$.

For reasons which will become apparent, an additional 90° phase-shift element is explicitly included. The representation of such a phase shift is

$$S_1 = \begin{pmatrix} 0 & j \\ j & 0 \end{pmatrix}$$

The symbols to be used are shown in Fig. 2, where the unprimed E is the input amplitude and the primed E is the

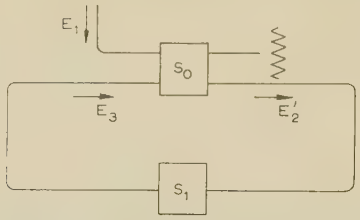


Fig. 2.—Simple travelling-wave cavity.

output amplitude, the subscripts referring to the particular arm of the directional coupler as defined in Fig. 1.

One gets simply

$$E_2' = aE_1 + jsE_3 \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

and since the amplitude E_2 travels the length, l, of the cavity,

$$E_3 = j\varepsilon^{j\phi}\eta E_2' \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Solving, one gets

$$\frac{E_2'}{E_1} = \frac{a}{1 + s\eta\varepsilon^{j\phi}} \quad . \quad . \quad . \quad . \quad . \quad (3)$$

The power ratio is

$$\left|\frac{E_2'}{E_1}\right|^2 = \frac{a^2}{1 + s^2\eta^2 + 2s\eta\cos\phi} \quad . \quad . \quad . \quad (4)$$

The resonance condition is $\varepsilon^{j\phi} = -1$, so that $\phi = (2n + 1)\pi$ where n is some integer.

At resonance,

$$\left|\frac{E_2'}{E_1}\right|^2 = \frac{a^2}{(1 - s\eta)^2} \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Off resonance, letting $\phi = (2n + 1)\pi + \delta$,

$$\left|\frac{E_2'}{E_1}\right|^2 = \frac{a^2}{1 + K^2 - 2K\cos\delta} \quad . \quad . \quad . \quad (6)$$

where K is the cavity constant defined by

$$K = s\eta \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

The ratio of the off-resonance to on-resonance power ratios [eqns. (5) and (6)] is defined as $P_0(\delta)$, so that

$$P_0(\delta) = \frac{(1 - K)^2}{1 + K^2 - 2K\cos\delta} \quad . \quad . \quad . \quad (8)$$

In practice, the cavity constant K may be determined by measuring the power ratio as the phase shift around the cavity is adjusted by a phase shifter in the loop. Solving eqn. (8) for K,

$$K = \lambda\left[1 - \sqrt{\left(1 - \frac{1}{\lambda^2}\right)}\right] \quad . \quad . \quad . \quad . \quad (9)$$

where

$$\lambda \equiv \frac{1 - P_0\cos\delta}{1 - P_0} \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

If an actual cavity fulfils the above requirements and is detuned by a given amount, δ, P_0 may be measured and the cavity constant may be computed by eqn. (9). The results of a series of such calculations using various δ's and the corresponding P_0's will give a constant value for K. Experimentally, a constant value was not obtained; rather, a systematic variation in the computed values for K was found [see Fig. 6(*a*)]. We consider now a possible factor for such a systematic dependence of the computed K on δ.

Cavity with a Small Internal Reflection (Perturbed Cavity).—Consider the effect of a small internal mismatch in the cavity. Such a cavity will be referred to as the perturbed cavity. A lossless reflector is represented symbolically by

$$S_2 = \begin{pmatrix} \rho & j\tau \\ j\tau & \rho \end{pmatrix}$$

where $\rho^2 + \tau^2 = 1$. Any loss in a reflector may be included in the characteristic loss, η, of the cavity. The symbols to be used are shown in Fig. 3, where the anticipated reflected amplitudes are inserted. The results here obtained reduce in the limiting case of $\rho = 0(\tau = 1)$ to those in the previous paragraph.

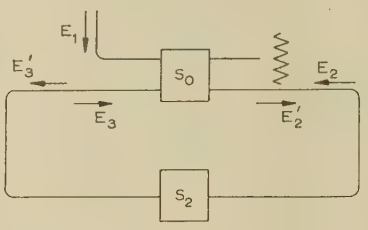


Fig. 3.—Travelling-wave cavity containing a small internal reflection.

One gets

$$E_2' = aE_1 + jsE_3 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

$$E_3' = jsE_2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

and

$$E_3 = j\tau\eta\varepsilon^{j\phi}E_2' + \rho\eta^2\varepsilon^{j2\phi}\varepsilon^{-j2\phi x/l}\varepsilon^{2\beta x}E_3' \quad . \quad (13)$$

$$E_2 = \rho\varepsilon^{j2\phi x/l}\varepsilon^{-\beta x}E_2' + j\tau\eta\varepsilon^{j\phi}E_3' \quad . \quad . \quad . \quad (14)$$

having assumed the reflector is at a distance x in the clockwise direction from S_0 (the directional coupler). The algebra yields

$$\frac{E_2'}{E_1} = \frac{a}{1 + \tau K\varepsilon^{j\phi} + \dfrac{\rho^2K^2\varepsilon^{j2\phi}}{1 + \tau K\varepsilon^{j\phi}}} \quad . \quad . \quad . \quad (15)$$

and using $\phi = (2n + 1)\pi + \delta$ and $D_0 \equiv 1 + \tau^2K^2 - 2\tau K\cos\delta$

$$\left|\frac{E_2'}{E_1}\right|^2 = \frac{a^2}{D_0\left[\left(1 + \dfrac{\rho^2K^2}{D_0}\right)^2 - \left(\dfrac{2\rho K\sin\delta}{D_0}\right)^2\right]} \quad . \quad (16)$$

We define a quantity $P_1(\delta)$ which is the ratio of eqn. (16) to the value of

$$\left|\frac{E_2'}{E_1}\right|^2$$

at resonance for the simple cavity [eqns. (5)]:

$$P_1(\delta) = \frac{(1 - K)^2}{D_0\left[\left(1 + \dfrac{\rho^2K^2}{D_0}\right)^2 - \left(\dfrac{2\rho K\sin\delta}{D_0}\right)^2\right]} \quad . \quad (17)$$

As stated previously, in the limit $\rho = 0(\tau = 1)$, $P_1(\delta) = P_0(\delta)$ [see eqn. (8)]. At resonance,

$$P_1(0) = \frac{(1 - K)^2}{(1 - \tau K)^2\left[1 + \dfrac{\rho^2K^2}{(1 - \tau K)^2}\right]^2} \quad . \quad . \quad (18)$$

which is less than unity since $\tau < 1$. This means that the power at resonance in a cavity which is mismatched will not be as great as in a perfectly matched cavity, even though the cavities have identical intrinsic losses. The result is expected because the reflector couples energy into a mode (the backward wave)

other than the one being observed, thus appearing to be a lossy element. The extent of this coupling increases as resonance is approached, so the effect of the reflector is to flatten out the cavity response near resonance. The greatest deviation from the response of the simple cavity is at resonance. One method of mimimizing the effect of the backward scattering is to insert a non-reciprocal element in the cavity which exhibits a large loss in the backward-wave direction. The Q-factor of the cavity for the backward-wave mode will then be lowered, with the result that less energy will be coupled into that mode.

Cavity with Internal Reflector and Isolator.—A perfectly matched isolator with zero forward loss is represented by the matrix

$$I = \begin{pmatrix} 0 & \alpha \\ 1 & 0 \end{pmatrix}$$

where $\alpha < 1$. The symbols to be used are shown in Fig. 4. Just as before, one has

$$E_2' = aE_1 + jsE_3 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

$$E_3' = jsE_2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

Further,

$$E_3 = j\tau\eta\varepsilon^{j\phi}E_2' + \rho\eta^2\varepsilon^{j2\phi}\varepsilon^{-j2\phi/l}\varepsilon^{2\beta x}E_3' \quad . \quad . \quad (21)$$

$$E_2 = \alpha\rho\varepsilon^{j2\phi x/l}\varepsilon^{-2\beta x}E_2' + j\tau\alpha\eta\varepsilon^{j\phi}E_3' \quad . \quad . \quad . \quad (22)$$

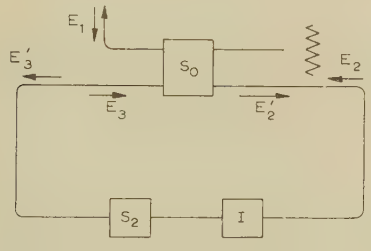


Fig. 4.—Travelling-wave cavity with an internal reflection and an isolator.

Eqns. (21) and (22) are slightly modified when the positions of I and S_2 are reversed; however, the resultant forward wave E_2' is unaffected (the magnitude of the backward wave is sensitive, by a factor of α, to the interchange). The result of solving for E_2' is

$$\frac{E_2'}{E_1} = \frac{a}{1 + \tau K\varepsilon^{j\phi} + \dfrac{\alpha\rho^2K^2\varepsilon^{j2\phi}}{1 + \alpha\tau K\varepsilon^{j\phi}}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (23)$$

$$\left|\frac{E_2'}{E_1}\right|^2 = \frac{a^2}{D_0\left\{\left(1 + \dfrac{\alpha\rho^2K^2}{D_0}\right)\left(1 + \dfrac{\alpha\rho^2K^2}{D_1}\right) - \dfrac{\alpha\rho^2K^2}{D_0D_1}[4\sin\delta + \tau^2K^2(1-\alpha)^2]\right\}} \quad . \quad . \quad . \quad . \quad . \quad (24)$$

where δ is again defined as the deviation from resonance, and $D_0 \equiv 1 + \tau^2K^2 - 2\tau K\cos\delta$; $D_1 \equiv 1 + \alpha^2\tau^2K^2 - 2\alpha\tau K\cos\delta$. Eqn. (24) properly reduces to eqn. (16) for $\alpha = 1$, since $D_1 = D_0$ when α is unity. We define $P_2(\delta)$ for this cavity as the ratio of eqn. (24) to eqn. (5), and so

$$P_2(\delta) = \frac{(1-K)^2}{D_0\left\{\left(1 + \dfrac{\alpha\rho^2K^2}{D_0}\right)\left(1 + \dfrac{\alpha\rho^2K^2}{D_1}\right) - \dfrac{\alpha\rho^2K^2}{D_0D_1}[4\sin^2\delta + \tau^2K^2(1-\alpha)^2]\right\}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (25)$$

At resonance,

$$P_2(0) = \frac{(1-K)^2}{(1-\tau K)^2\left\{\left[1 + \dfrac{\alpha\rho^2K^2}{(1-\tau K)^2}\right]\left[1 + \dfrac{\alpha\rho^2K^2}{(1-\alpha\tau K)^2}\right] - \dfrac{\alpha\rho^2K^2}{(1-\tau K)^2(1-\alpha\tau K)^2}[\tau^2K^2(1-\alpha)^2]\right\}} \quad . \quad . \quad (26)$$

A careful comparison of eqn. (26) with eqn. (18) shows that $P_2(0) > P_1(0)$ since $0 < \alpha < 1$. This means that the perturbed cavity with the isolator behaves more like the simple cavity than the perturbed cavity alone.

The Experiment

The schematic of the system used in the experiment is shown in Fig. 5. The tunable transformer in the loop was used to

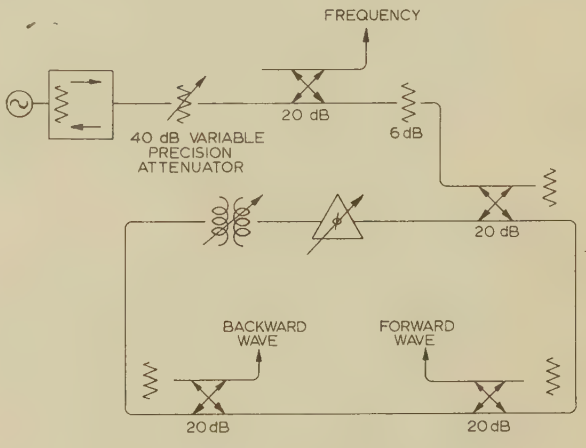


Fig. 5.—Schematic of experimental cavity.

tune out the backward wave as observed at the output of the backward-wave directional coupler. As the variable phase shifter was varied off resonance, the backward wave remained below the noise level. The variable precision attenuator was adjusted at each setting of the phase shifter to maintain a constant power level in the loop. Fig. 6(*a*) shows the calculated values for λ using the simple cavity equation. The fact that the computed values do not lie on a straight line indicates that other factors must be considered. The monitored backward wave was always below the noise level, so it may seem at first that a reflection in the cavity cannot be the other factor which must be considered. However, it will be shown that a very small reflection (below the noise level) is sufficient to explain the observed response.

Analysis of Experiment.—We have seen that a reflector in the

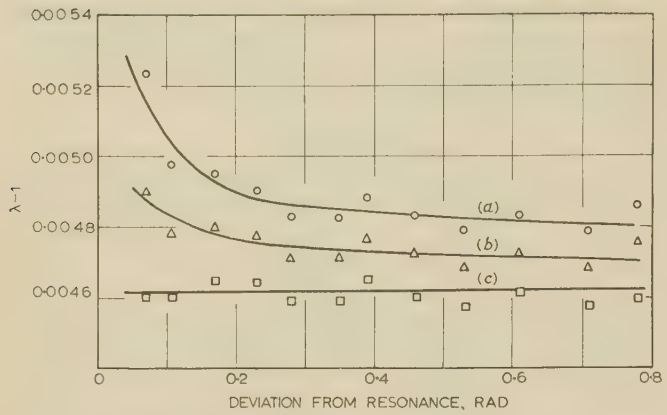


Fig. 6.—$\lambda - 1$ as a function of δ.

(*a*) Computed from simple cavity equation.
(*b*) Computed similarly, but assuming resonance reading 0·1 dB too low.
(*c*) Computed assuming resonance reading 0·2 dB too low.

cavity reduces the peak power in the cavity at resonance. We assume that this reduction was 0·1 dB and recompute λ on this basis. The resulting values for λ are shown in Fig. 6(*b*). Fig. 6(*c*) shows the λ calculations based on the assumption of a 0·2 dB error in power at resonance. The averaged value for $\lambda - 1$ on this latter assumption is 0·004 614 with a 0·8% discrepancy. Since this latter assumption and consequent calculation yielded the most consistent values for λ, we assume the above value for λ is the intrinsic value for the cavity. The cavity constant computed by eqn. (8) is $K = 0{\cdot}908\,441$. The question now is: How large a reflection is needed to produce a 0·2 dB reduction in the peak power for the cavity with K equal to the above value? Using the defining equations for P_0 and P_1 [eqns. (8) and (17)] the ratio P_1/P_0 at resonance is

$$R \equiv P_1/P_0 = \frac{(1-K)^2(1-\tau K)^2}{[(1-\tau K)^2 + \rho^2 K^2]^2} \quad . \quad . \quad . \quad (27)$$

The computed values for this ratio as a function of ρ, using the above given value for K, are plotted in Fig. 7. The value of this ratio corresponding to the 0·2 dB low reading for the peak is 0·955. From Fig. 7 we see that a ρ of 0·015 will give the 0·2 dB drop of the power at resonance. A ρ of 0·015 in turn

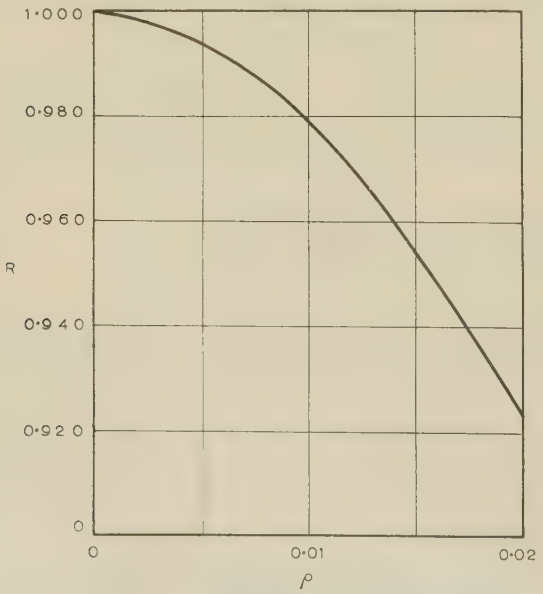


Fig. 7.—R as a function of ρ with $\lambda - 1 = 0{\cdot}004\,614$.

corresponds to a discontinuity reflecting only −36·5 dB of the incident power.

Using $\rho = 0{\cdot}015$ and $K = 0{\cdot}908\,441$, the theoretical response P_1 of a cavity was computed, and the comparison of the theoretical curve with the actual response of the cavity is shown in Fig. 8. The curve was fitted at resonance. It is seen that

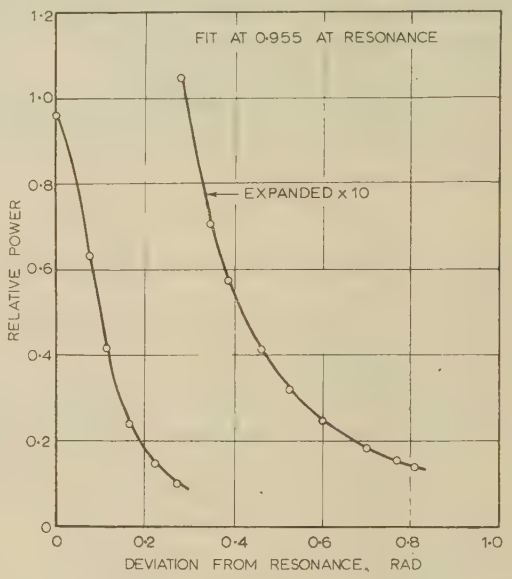


Fig. 8.—Relative power in cavity as a function of the deviation from resonance.

○ Experimental points.
—— Theoretical data.

the experimental points fall almost exactly on the theoretical curve. For a more critical comparison the theoretical response curve was treated as if the simple cavity relationships held, and the λ's for various points were computed. This theoretical curve is plotted in Fig. 9, as are the similarly computed experimental points which were plotted previously in Fig. 6(*a*).

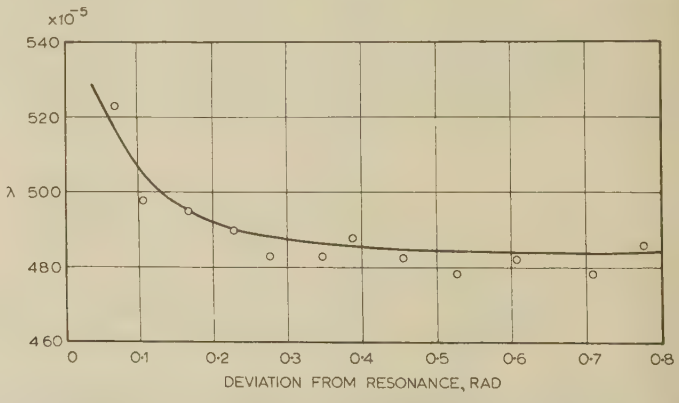


Fig. 9.—λ as a function of the deviation from resonance.

○ Experimental points.
—— Theoretical data.
λ (true) = 0·004 614 (K = 0·908 441).

As interesting as the large adverse effect on the response of the cavity produced by a small reflector is the counter-effect produced by an isolator inserted in the cavity. The theoretical response of the same cavity but with an isolator of 20 dB backward attenuation ($\alpha = 0{\cdot}1$) was computed by eqn. (25) and plotted in Fig. 10, together with the perturbed and simple cavity responses. The resultant response differs at the peak by

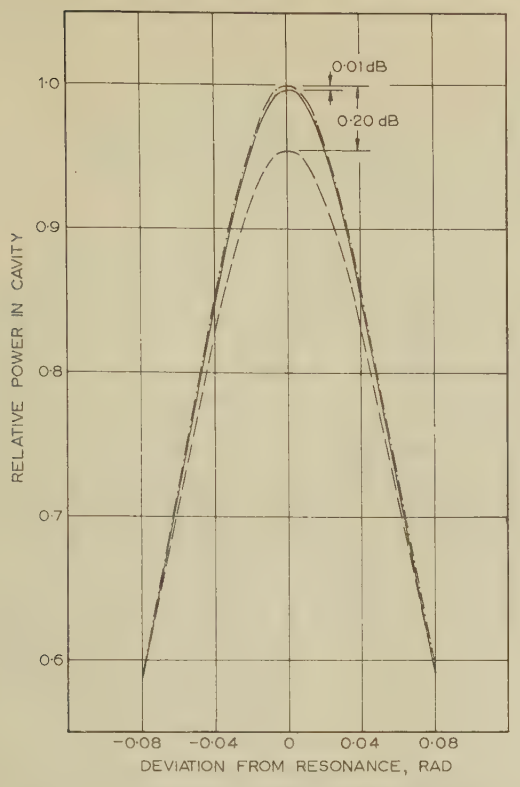


Fig. 10.—Relative power in the cavity as a function of the deviation from resonance for the three cavities considered.

-·-·- Simple cavity response.
——— With −36 dB reflection and 20 dB isolator.
- - - - With −36 dB reflection and no isolator.

only 0·01 dB from the simple cavity response. An experimental check was not possible because no available isolator had a small enough forward loss to correspond to the theoretically assumed isolator.

Conclusions

The agreement of the theory with the experiment gives one confidence in the predicted response of the cavity with an isolator. Although a physically realizable isolator will have a certain amount of forward loss which will increase the intrinsic loss of the cavity, the advantage of including the isolator is that the cavity will respond very much as the simple cavity. The cavity, therefore, can be analysed using the relatively simple equations (6), (7) and (8). The result is a very happy one because the prospects of having to use eqn. (17) to interpret his result would discourage many a researcher.

References

(1) Tischer, F. J., Lindberg, W. J., and Copeland, J.: 'Resonance Properties of Ring Circuits', *Transactions of the Institute of Radio Engineers*, 1957, **MTT–5**, p. 51.

(2) Sferrazza, P. J.: 'Traveling Wave Resonator', *Tele-Tech and Electrical Industries*, November, 1955, **14**, p. 84.

(3) Milosevic, L.: 'Résonateur à ondes progressives, I: Généralités et principe de fonctionnement', *Revue Technique de la Compagnie Française de Thomson-Houston*, 1955, No. 21, p. 61.

(4) Ault, L. A., Spencer, E. G., and Le Craw, R. C.: 'Circularly Polarized Traveling Wave Cavity for Ferrite Tensor Permeability Measurements', Diamond Ordnance Fuze Laboratories, September, 1957, TR–482.

(5) Tomiyasu, K.: 'Effect of a Mismatched Ring in a Traveling-Wave Resonant Circuit', *Transactions of the Institute of Radio Engineers*, 1957, **MTT–5**, p. 267.

621.373.14

The Institution of Electrical Engineers
Paper No. 3684 E
Sept. 1961

AN ANALYSIS OF A CYLINDRICAL CAVITY WITH RADIAL VANES

By A. SINGH, Ph.D., and R. A. RAO.

(*The paper was first received* 19*th November*, 1960, *and in revised form* 2*nd June*, 1961.)

SUMMARY

An analysis is given for evaluating the resonance frequencies of modes of lower orders in a cylindrical cavity having radial vanes. The procedure involves the use of variational methods for obtaining successively better approximations to the resonance frequencies and the field configurations. The trial field is chosen to satisfy the boundary conditions at the vanes. This makes the procedure valid even when the perturbations introduced by the vanes are large. Thus good agreement is obtained between computed and experimentally observed resonance frequencies for varying degrees of penetration by a pair of vanes, the change in frequency produced by the vanes being up to 28%.

LIST OF SYMBOLS

Rationalized M.K.S. units are employed.

a = Distance from axis of cavity to tip of vanes.
b = Radius of cavity.
d = Radial length of the vanes.
TM_{n10} = Transverse magnetic mode with azimuthal order n, radial order one and axial order zero.
ϕ, r, z = Azimuthal, radial and axial co-ordinates.
$E_z(r, \phi), H_\phi(r, \phi)$ = Distributions of E_z and H_ϕ as functions of r, ϕ. Additional subscript 1 or 2 refers to region 1 or 2 defined in Fig. 1.
A_ν, B_ν = Fourier components of $E_{z1}(a, \phi)$ and $E_{z2}(a, \phi)$.
$\beta = 2\pi f/c$, where f is the frequency and c the velocity of light in vacuum.
$J_n(x)$ = Bessel function of first kind, nth order and argument x.
$N_n(x)$ = Bessel function of second kind, nth order and argument x.

$$Q_{2\nu}(\beta a, \beta b) = J_{2\nu}(\beta a)N_{2\nu}(\beta b) - N_{2\nu}(\beta a)J_{2\nu}(\beta b).$$

$$Q'_{2\nu}(\beta a, \beta b) = \left\{\frac{\partial}{\partial(\beta r)}[J_{2\nu}(\beta r)N_{2\nu}(\beta b) - N_{2\nu}(\beta r)J_{2\nu}(\beta b)]\right\}_{r=a}.$$

(1) INTRODUCTION

In a study of mode control of an inter-digital magnetron[1] a method involving the use of radial vanes was found to give promising results. Vanes with a controlled degree of penetration were used to locate the modes of the lowest frequencies at desired intervals from one another. These modes are derived from the TM_{n10} modes of a simple cylindrical resonator. In order to compute the frequency spectrum of these modes, the analysis of the inter-digital resonator with radial vanes was taken up in two stages.*

The first stage was an analysis of a simple cylindrical resonator with radial vanes, as presented in the paper. In the second stage, which will be described in a subsequent paper, the analysis was extended to an inter-digital resonator with radial vanes. Here also good agreement between calculated and experimental results was obtained.

Even the first stage of the problem is rather difficult to handle using perturbation methods discussed in the literature.[2, 3] These methods employ trial fields which, together with their first derivatives, are continuous everywhere in the cavity (except at surfaces where ϵ and μ are discontinuous). Trial fields of this type, which also satisfy the boundary conditions at the perturbed surface, would be difficult to formulate. In order to surmount this difficulty, a category of trial fields can be employed which do not satisfy the boundary conditions at the perturbed surface. In this case, the computations involve an integral over the volume enclosed between the perturbed and unperturbed surfaces (or a surface integral that reduces to a similar volume integral). The volume thus enclosed, in the case of vanes assumed to be infinitesimally thin, is also infinitesimally small. Thus the effect of this type of perturbation cannot be accurately evaluated by using trial fields that do not satisfy the boundary conditions on the surface of the vanes.

In the following, an alternative approach has been used in tackling the problem of vanes. Briefly, it involves the following. The resonator is appropriately divided into separate regions, for which solutions to Maxwell's equations can be found such that they satisfy the boundary conditions on the walls. Next it is necessary to match the fields across the dividing surface. To match them at every point would again be a formidable task. However, a method of matching has been worked out such that quite accurate resonance frequencies can be obtained without resorting to exact matching at every point. To begin with, a physically significant parameter (the Poynting vector) integrated over the entire dividing surface is matched across it. In this way a good approximation to the frequency is obtained, even though the approximation to the field distribution at the dividing surface may not be very good. Next, the condition of point-by-point matching is reduced to a variational principle, which helps in improving upon the choice of the field distribution assumed at the dividing surface. By using these two steps alternately, one can get successively better approximations to the resonance frequency.

The same general approach has been used with success for the case of a disc-loaded circular waveguide.[4–8] However, in that case, the choice of trial fields is relatively simple in comparison with the case of radial vanes considered here. A detailed survey of the essential features of different variational methods has been given by Chu.[9]

The salient steps in the analysis of the present case are explained in Section 2. The analysis is next given in more detail for the TM_{010} and TM_{110} modes. Finally, a comparison of computed and experimental results is given, showing good agreement between the two. Major steps in the derivations for the case of the TM_{010} mode are given in Section 9.

(2) OUTLINE OF THE PROCEDURE

The analysis proceeds by dividing the cavity into different regions by an assumed cylindrical surface passing through the

* Preliminary work on this problem was done by Dr. Singh at the Cruft Laboratory, Harvard University, U.S.A. The work was continued at the Central Electronics Engineering Research Institute, Pilani, India.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Dr. Singh is Deputy Director-in-Charge of, and Mr. Rao was formerly at, the Central Electronics Engineering Research Institute, Pilani, India. Mr. Rao is now in the Department of Electrical Engineering, University of California, U.S.A.

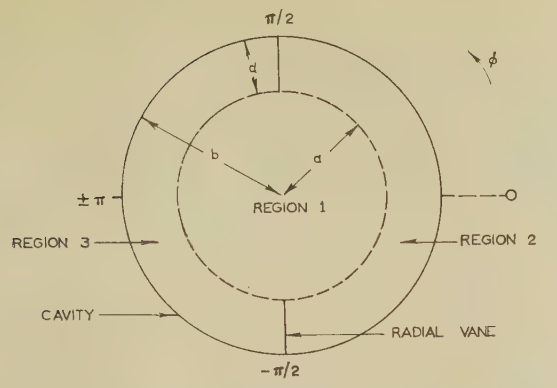


Fig. 1.—Cavity and vanes, illustrating the method of analysis.

inner edges of the vanes (Fig. 1). Electromagnetic fields are built up in regions 1, 2 and 3, such that they satisfy the boundary conditions at the walls of the cavity and the vanes, and Maxwell's equations within the given regions, and add up to an assumed distribution for either the *E*- or *H*-field at the dividing surface. The process of matching the *H*- or *E*-field across the dividing surface yields the correct field distribution and frequency. This matching is initially done for regions 1 and 2 (shown in Fig. 1) independently of region 3. It follows for regions 1 and 3 from symmetry conditions.

The fields satisfying the above requirements are built up by a summation of the field distribution for orthogonal modes of a simple cylindrical cavity, with coefficients decided by a Fourier analysis of the assumed distribution at the dividing surface. It is to be noted that in these distributions the frequency is also involved.

Initially, let the distribution be assumed for the *E*-field. For region 1 the Fourier analysis is carried out for a function which is equal to the assumed distribution over the entire range of ϕ from $-\pi$ to π. However, for region 2, a different function is chosen for analysis, which agrees with the above distribution in the range—$\pi/2$ to $\pi/2$, but is antisymmetric about the points $\phi = \pm \pi/2$, representing the azimuthal location of the vanes. The orthogonal modes which are thus obtained from a Fourier analysis of the function for region 2 all have a zero *E*-field at $\phi = \pm \pi/2$, not only at $r = a$ but also for $a < r < b$, i.e. for the entire surface of the vanes. Thus the boundary condition is satisfied on the vanes.

The matching of fields proceeds in successive iterative steps as follows. First, an approximation to the frequency is obtained from an equation which arises out of matching the surface integral of the Poynting vector across the dividing surface between regions 1 and 2.

This step is justifiable, since the phenomenon of resonance essentially implies the build-up of relatively large amplitudes of oscillation with a small supply of power from outside (an infinitesimally small supply in the case of perfectly conducting walls and infinitesimally small coupling). Consequently, if the resonator is considered to be made of two parts, the requirement for resonance is that the rates of total transfer of energy across the dividing surface are equal and opposite, so that the resonator as a whole may be self-sufficient. Thus, if we start with a certain field distribution at the dividing surface, then by using a physically significant integral over this surface, we can obtain a reasonably good value of the frequency even though the continuity conditions may not be satisfied point by point.

However, the exact solution demands such a point-by-point continuity. This condition was used to derive a variational principle for the field distribution, and thus represents a convenient way of obtaining a more exact field distribution. This variational principle holds rigorously when the correct value of frequency is inserted into the expressions. However, in practice, by using the above steps alternately, successively more accurate values of frequency and field distribution are obtained.

The same process can be carried out by starting with an *H*-field distribution at the dividing surface. In this case, the correct frequency is approached from the opposite direction to that obtained in the previous case, under conditions which will be clear in the specific examples given in the following Section. The limits within which the frequency lies can then be ascertained.

(3) TM_{010} MODE PERTURBED BY TWO VANES

In the case of the TM_{010} mode perturbed by two vanes, the *E*-field distribution at the dividing surface will have the form shown in Figs. 2(*a*) and (*b*), which also indicate the types of

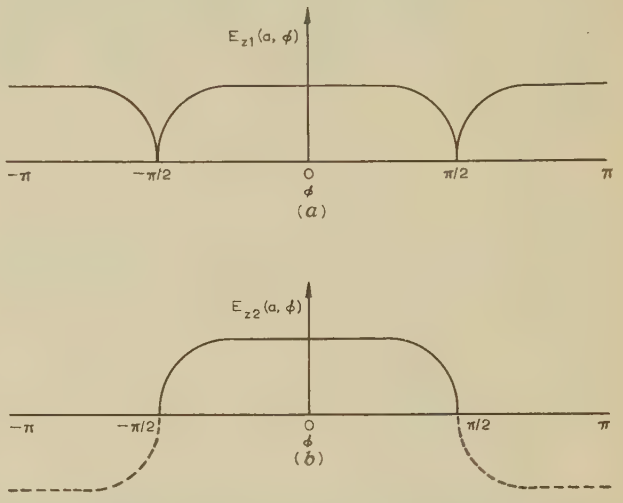


Fig. 2.—Type of *E*-field distribution analysed for perturbed TM_{010} mode.
(*a*) Region 1 at $r = a$.
(*b*) Region 2 at $r = a$.

function to be Fourier analysed for regions 1 and 2. By following the steps given in Section 9.1, the surface integral of the Poynting vector can be evaluated separately for the two regions over the surface dividing them from each other. Equating these integrals, integrating term-wise and simplifying,

$$\frac{J_0'(\beta a)}{J_0(\beta a)} = \frac{2}{\left[\int_{-\pi/2}^{\pi/2} E_z(a, \phi) d\phi\right]^2} \times \left\{ \sum_{\nu=0}^{\infty} \left[\int_{-\pi/2}^{\pi/2} E_z(a, \phi) \cos(2\nu + 1)\phi d\phi\right]^2 \frac{Q_{2\nu+1}'(\beta a, \beta b)}{Q_{2\nu+1}(\beta a, \beta b)} - \sum_{\nu=1}^{\infty} \left[\int_{-\pi/2}^{\pi/2} E_z(a, \phi) \cos 2\nu\phi d\phi\right]^2 \frac{J_{2\nu}'(\beta a)}{J_{2\nu}(\beta a)} \right\} \quad . \quad (1)$$

If $E_z(a, \phi)$ is the correct field distribution, the solution of eqn. (1) will give the correct value of β, and therefore of f, the resonance frequency.

The next step is to examine the variations in the right-hand side of eqn. (1) when β is assumed to be correct and $E_z(a, \phi)$ is varied from the correct distribution. For convenience the right-hand side of eqn. (1) is denoted by Y for the assumed field distribution $E_z(a, \phi)$, and is denoted by Y^0 for the correct field distribution $E_z^0(a, \phi)$.

Let $E_z(a, \phi) = E_z^0(a, \phi) + \Delta E_z(a, \phi)$, where $\Delta E_z(a, \phi)$ represents small variations from the correct distribution.

Also let $Y = Y^0 + \Delta Y$.

It is shown in Section 9.2 that under the conditions $H_{\phi 1}(a, \phi) = H_{\phi 2}(a, \phi)$ and β correct, ΔY is given by eqn. (2), omitting the writing of (a, ϕ) in the terms such as $E_z(a, \phi)$, $\Delta E_z(a, \phi)$, etc.

$$\Delta Y = \frac{1}{\left(\int_{-\pi/2}^{\pi/2} E_z d\phi\right)^2} \left\{ \sum_{\nu=0}^{\infty} \left[\int_{-\pi/2}^{\pi/2} \Delta E_z \cos(2\nu+1)\phi d\phi \right]^2 \times \right.$$

$$\left. \frac{Q'_{2\nu+1}(\beta a, \beta b)}{Q_{2\nu+1}(\beta a, \beta b)} - \sum_{\nu=1}^{\infty} \left[\int_{-\pi/2}^{\pi/2} \Delta E_z \cos 2\nu\phi d\phi \right]^2 \frac{J'_{2\nu}(\beta a)}{J_{2\nu}(\beta a)} \right\} \quad . \quad (2)$$

Eqn (2) shows that ΔY has its sign unchanged if $\Delta E_z(a, \phi)$ is replaced by $-\Delta E_z(a, \phi)$, where ΔE is assumed to be small compared with E_z^0. Hence Y is stationary at the correct field distribution, with respect to small variations in the field distribution. By adjusting the Fourier components of the assumed field distribution to make Y stationary, one can work towards the equality of $H_{\phi 1}(a, \phi)$ and $H_{\phi 2}(a, \phi)$, and thus get a closer approximation to the correct field distribution. It is to be noticed that since

$$H_r = \frac{1}{\beta^2} \frac{j\omega\epsilon}{r} \frac{\partial E_z}{\partial \phi}$$

the continuity condition for H_r across the dividing surface will also be satisfied if it is satisfied for E_z at all values of ϕ in the range $-\pi/2$ to $\pi/2$. Also $J'_0(\beta a)/J(\beta a)$ has a negative sign for $0 < \beta a < 2{\cdot}40$, which range covers all possibilities in the case of the TM_{010} mode.

Thus if $$\frac{J'_{2\nu}(\beta a)}{J_{2\nu}(\beta a)} \geqslant 0 \qquad \text{for } \nu = 1, 2, \ldots$$

and $$\frac{Q'_{2\nu+1}(\beta a, \beta b)}{Q_{2\nu+1}(\beta a, \beta b)} \leqslant 0 \qquad \text{for } \nu = 0, 1, 2 \ldots$$

ΔY will have the same sign as Y regardless of the choice of ΔE_z. In such a case we can also say that $|Y^0| \leqslant |Y|$ for all trial fields.

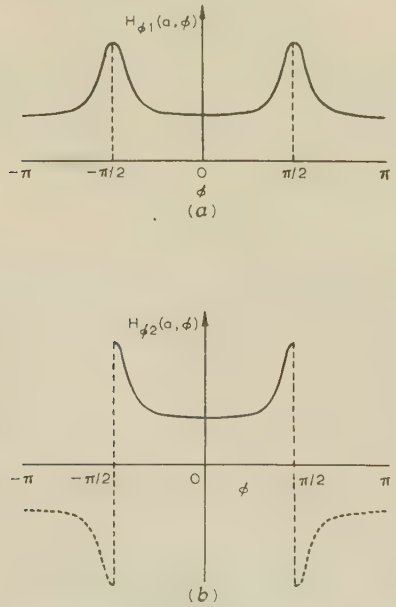


Fig. 3.—Type of H-field distribution analysed for perturbed TM_{010} mode.

(*a*) Region 1 at $r = a$.
(*b*) Region 2 at $r = a$.

The same kind of analysis can be carried out by starting with an assumed distribution of $H_\phi(a, \phi)$ instead of $E_z(a, \phi)$. In this case the general form of the functions, which are Fourier analysed to obtain the field distribution in regions 1 and 2, is shown in Figs. 3(*a*) and (*b*). Again the boundary condition for H_ϕ is exactly satisfied at the vanes. By following similar steps, the following equation is obtained in place of eqn. (1):

$$\frac{J_0(\beta a)}{J'_0(\beta a)} = \frac{2}{\left[\int_{-\pi/2}^{\pi/2} H_\phi(a, \phi) d\phi\right]^2} \left\{ \sum_{\nu=0}^{\infty} \left[\int_{-\pi/2}^{\pi/2} H_\phi(a, \phi) \cos(2\nu+1)\phi d\phi \right]^2 \times \right.$$

$$\left. \frac{Q_{2\nu+1}(\beta a, \beta b)}{Q'_{2\nu+1}(\beta a, \beta b)} - \sum_{\nu=1}^{\infty} \left[\int_{-\pi/2}^{\pi/2} H_\phi(a, \phi) \cos 2\nu\phi d\phi \right]^2 \frac{J_{2\nu}(\beta a)}{J'_{2\nu}(\beta a)} \right\} \quad . \quad (3)$$

The right-hand side of eqn. (3) is denoted by Z for the assumed distribution $H_\phi(a, \phi)$, and by Z^0 for the correct distribution $H_\phi^0(a, \phi)$. By proceeding in an analogous way to the previous case, it can be shown that Z is stationary about the correct field distribution. Thus a similar iteration process can be carried out. It is also seen from eqns. (1) and (3) that, for the correct field distribution, $Y^0 = 1/Z^0$.

Again it can be shown that if

$$\frac{Q_{2\nu+1}(\beta a, \beta b)}{Q'_{2\nu+1}(\beta a, \beta b)} \leqslant 0 \qquad \text{for } \nu = 0, 1, 2 \ldots$$

and $$\frac{J_{2\nu}(Ba)}{J'_{2\nu}(\beta a)} \geqslant 0 \qquad \text{for } \nu = 1, 2 \ldots$$

(which are essentially the same conditions as before), ΔZ will have the same sign as Z, so that $|Z^0| \leqslant |Z|$. In such a case the use of both the methods gives upper and lower limits, respectively, for $J'_0(\beta a)/J_0(\beta a)$ for the two cases. This enables an upper and lower value for β to be determined also.

(4) TM_{110} MODE PERTURBED BY TWO VANES AT THE MAXIMA OF THE *E*-FIELD

The method of solution for the TM_{110} mode perturbed by two vanes at the maxima of the E-field is exactly the same as before, except that the functions approximated by the Fourier series, in order to obtain the field distribution in regions 1 and 2, are now different. They are shown for the case of the E-field in Figs. 4(*a*) and (*b*).

The equations obtained corresponding to eqns. (1) and (3) are

$$\frac{J'_1(\beta a)}{J_1(\beta a)} = \frac{1}{\left[\int_{-\pi/2}^{\pi/2} E_z(a, \phi) \sin\phi d\phi\right]^2} \left\{ \sum_{\nu=1}^{\infty} \left[\int_{-\pi/2}^{\pi/2} E_z(a, \phi) \sin 2\nu\phi d\phi \right]^2 \times \right.$$

$$\left. \frac{Q'_{2\nu}(\beta a, \beta b)}{Q_{2\nu}(\beta a, \beta b)} - \sum_{\nu=1}^{\infty} \left[\int_{-\pi/2}^{\pi/2} E_z(a, \phi) \sin(2\nu+1)\phi d\phi \right]^2 \frac{J'_{2\nu+1}(\beta a)}{J_{2\nu+1}(\beta a)} \right\} \quad . \quad (4)$$

$$\frac{J_1(\beta a)}{J'_1(\beta a)} = \frac{1}{\left[\int_{-\pi/2}^{\pi/2} H_\phi(a, \phi) \sin\phi d\phi\right]^2} \left\{ \sum_{\nu=1}^{\infty} \left[\int_{-\pi/2}^{\pi/2} H_\phi(a, \phi) \sin 2\nu\phi d\phi \right]^2 \times \right.$$

$$\left. \frac{Q_{2\nu}(\beta a, \beta b)}{Q'_{2\nu}(\beta a, \beta b)} - \sum_{\nu=1}^{\infty} \left[\int_{-\pi/2}^{\pi/2} H_\phi(a, \phi) \sin(2\nu+1)\phi d\phi \right]^2 \frac{J_{2\nu+1}(\beta a)}{J'_{2\nu+1}(\beta a)} \right\} \quad . \quad (5)$$

(5) COMPUTATIONS AND COMPARISON WITH EXPERIMENTAL RESULTS

Calculations were made and verified for the TM_{010} and TM_{110} modes in a cavity of radius $b = 1{\cdot}500$ in with pairs of

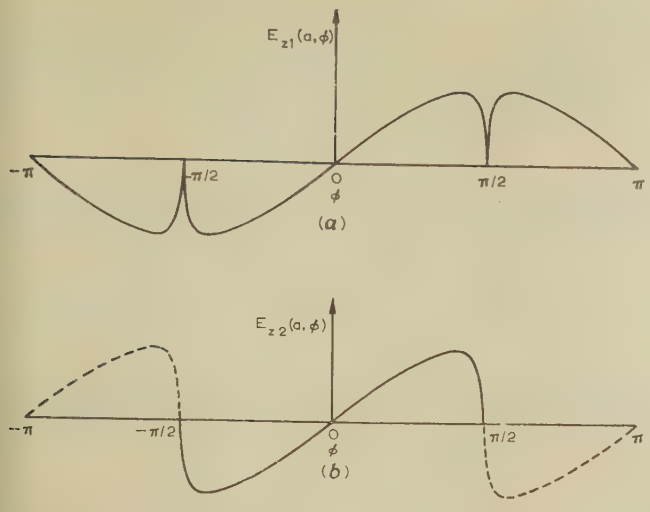


Fig. 4.—Type of *E*-field distribution analysed for perturbed TM_{110} mode.

(*a*) Region 1 at $r = a$.
(*b*) Region 2 at $r = a$.

vanes of radial length $d = 0{\cdot}300$, $0{\cdot}600$, $0{\cdot}750$ and $0{\cdot}900$ in. An outline of the computation for the TM_{010} mode is as follows.

Taking only one parameter for adjustments in the trial fields, the type of the distribution indicated in Figs. 2(*a*) and (*b*) can be obtained if one chooses $E_{z2}(a, \phi)$ of the form $(\cos\phi - \alpha_E \cos 3\phi)$. Similarly the type of distribution indicated in Figs. 3(*a*) and (*b*) can be obtained if one chooses $H_{\phi 1}(a, \phi)$ of the form $(1 - \alpha_H \cos 2\phi)$. α_E and α_H are the parameters to be adjusted. In the two cases, $E_{z1}(a, \phi)$ and $H_{\phi 2}(a, \phi)$ are derived by reflecting the previous functions antisymmetrically about $\phi = -\pi/2$ and $\pi/2$.

Eqns. (1) and (3) can be solved graphically. The solutions can be conveniently obtained from an initial plot of $J_0'(\beta a)/J_0(\beta a)$ against βa. This is illustrated in Fig. 5 for the case $a/b = 0{\cdot}4$. The conditions for $|Y^0|$ to be less than or equal to $|Y|$ and for $|Z^0|$ to be less than or equal to $|Z|$ are satisfied. In successive iterative steps, eqn. (1) gives βa as $1{\cdot}33$, $1{\cdot}235$ and $1{\cdot}230$, corresponding to $\alpha_E = 0{\cdot}4$, $0{\cdot}154$ and $0{\cdot}13$. Eqn. (3) gives βa as $1{\cdot}17$, $1{\cdot}185$ and $1{\cdot}185$ corresponding to $\alpha_H = 1{\cdot}0$, $0{\cdot}819$ and $0{\cdot}738$. The mean of the last two values of βa in the two sets is $1{\cdot}208$ and corresponds to a frequency of 3 783 Mc/s, which was taken as the computed value. The experimental value in this case was 3 810 Mc/s.

Considering all the computed and experimental values, it was found that they agreed to within a mean deviation of $0{\cdot}3\%$ for the TM_{010} mode and $0{\cdot}6\%$ for the TM_{110} mode. The maximum changes between perturbed and unperturbed frequencies were 26% and 28%, respectively, in the two modes.

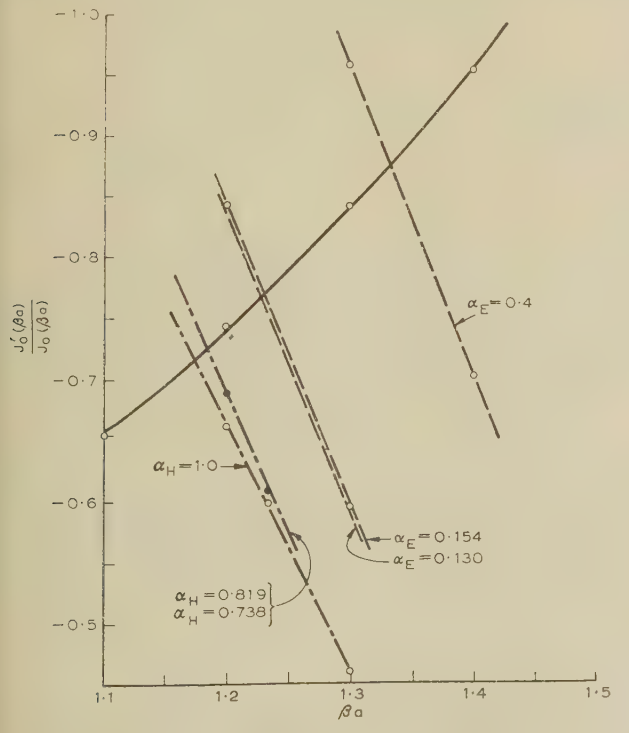


Fig. 5.—Evaluation of perturbed resonance frequency of TM_{010} mode for $a/b = 0{\cdot}4$, in iterative steps (in terms of βa).

——— $J_0'(\beta a)/J_0(\beta a)$.
– – – Y.
—·—·— $1/Z$.

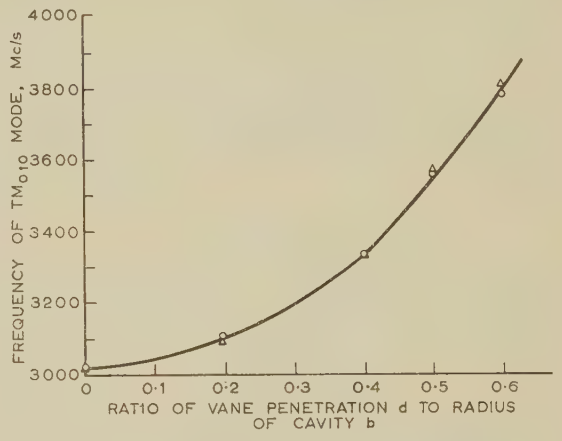


Fig. 6.—Variation of computed and observed frequency of TM_{010} mode with radial penetration by a pair of vanes.

○ Calculated.
△ Observed.

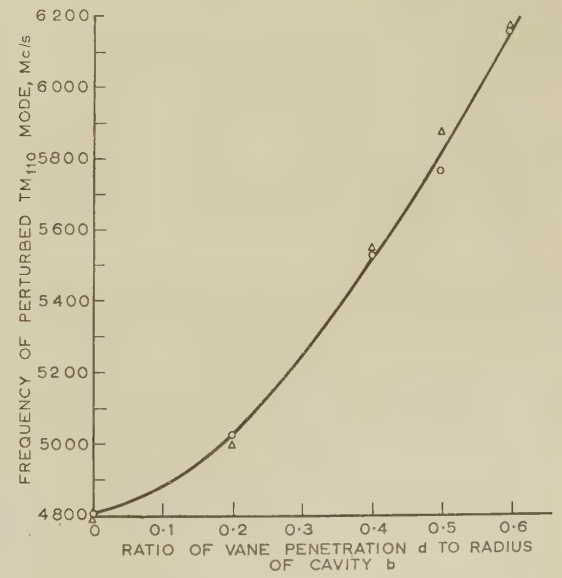


Fig. 7.—Variation of computed and observed frequency of TM_{110} mode with radial penetration by a pair of vanes.

○ Calculated.
△ Observed.

These results are plotted in Figs. 6 and 7. It is seen that the approach is applicable with good accuracy to even large perturbations by vanes. Still greater accuracy may be obtained by using more than one variable parameter to specify the trial field.

(6) CONCLUSIONS

An analysis has been given for evaluating resonance frequencies of the lower-order modes of a cylindrical cavity perturbed by radial vanes. The trial fields satisfy the boundary conditions at the vanes. An iterative procedure has been given for matching of fields across an assumed dividing surface. A variational principle used in the above iteration has been derived.

The method has been verified for varying degrees of vane penetration by a pair of vanes perturbing the TM_{010} and TM_{110} modes. The computed and experimental values show good agreement up to frequency changes of 28% between perturbed and unperturbed values.

This type of approach may also be useful in other cases where large perturbations are involved.

(7) ACKNOWLEDGMENTS

Many helpful discussions with Dr. W. S. Lucke in the early stages of the work, and the help of Messrs. G. S. Sidhu and H. S. Dewan in the experimental observations, are thankfully acknowledged.

(8) REFERENCES

(1) SINGH, A.: 'Modes and Operating Voltages of Interdigital Magnetrons', *Proceedings of the Institute of Radio Engineers*, 1955, **43**, p. 470.
(2) SLATER, J. C.: 'Microwave Electronics' (Van Nostrand, 1950), p. 81.
(3) BERK, A. D.: 'Variational Principles for Electromagnetic Resonators and Waveguides', *Transactions of the Institute of Radio Engineers*, 1956, **AP-4**, p. 104.
(4) CHU, E. L., and HANSEN, W. W.: 'Disk-Loaded Waveguides', *Journal of Applied Physics*, 1949, **20**, p. 280.
(5) WALKINSHAW, W., and BELL, J. S.: A.E.R.E. Report G/R 675, 1951.
(6) BELL, J. S.: A.E.R.E. Report G/R 680, 1951.
(7) WALKINSHAW, W., and BELL, J. S.: A.E.R.E. Report T/R 864, 1952.
(8) COLLIN, R. E.: 'Field Theory of Guided Waves' (McGraw-Hill, 1960).
(9) CHU, E. L.: 'Upper and Lower Bounds of Eigenvalues for Composite-Type Regions', *Journal of Applied Physics*, 1950, **21**, p. 454.

(9) APPENDICES: DERIVATIONS FOR TM_{010} MODE

(9.1) Expression Resulting from Equating Surface Integrals of Poynting Vector

The field distribution of the normal modes of the category TM_{n10} in a cylindrical cavity is given by

$$E_z = [p_\nu J_\nu(\beta r) + q_\nu N_\nu(\beta r)](c_\nu \cos \nu\phi + d_\nu \sin \nu\phi) \quad . \quad (6)$$

$$H_r = -\frac{1}{\beta^2}\frac{j\omega\epsilon}{r}\frac{\partial E_z}{\partial \phi} \quad . \quad (7)$$

$$H_\phi = -\frac{1}{\beta^2} j\omega\epsilon \frac{\partial E_z}{\partial r} \quad . \quad (8)$$

$$H_z = E_r = E_\phi = 0 \quad . \quad (9)$$

where p_ν, q_ν, c_ν and d_ν are constants. Let an assumed distribution of E_z at the dividing surface be denoted by $E_z(a, \phi)$, and the functions to be Fourier analysed for the regions 1 and 2, as illustrated in Fig. 2, be denoted by $E_{z1}(a, \phi)$ and $E_{z2}(a, \phi)$ respectively. As outlined in the main text,

$$E_{z1}(a, \phi) = E_z(a, \phi) \text{ for } -\pi < \phi < \pi.$$

Since $E_z(a, \phi)$ is symmetrical about $\phi = \pm \pi/2$, we can write $E_{z2}(a, \phi)$ as follows, making it antisymmetrical about $\phi = \pm \pi/2$:

$$E_{z2}(a, \phi) = E_z(a, \phi) \quad \text{for } -\pi/2 < \phi < \pi/2$$
$$\text{and } E_{z2}(a, \phi) = -E_z(a, \phi) \quad \text{for } -\pi < \phi < -\pi/2 \text{ and } \pi/2 < \phi < \pi.$$

Noticing that $E_{z1}(a, \phi)$ and $E_{z2}(a, \phi)$ are also symmetrical about $\phi = 0$ and $\pm\pi$, the Fourier expansions take the following form:

$$E_{z1}(a, \phi) = \sum_{\nu=0}^{\infty} A_{2\nu} \cos 2\nu\phi \quad . \quad (10)$$

where
$$A_0 = \frac{1}{\pi}\int_{-\pi/2}^{\pi/2} E_z(a, \phi)d\phi \quad . \quad (11a)$$

and
$$A_{2\nu} = \frac{2}{\pi}\int_{-\pi/2}^{\pi/2} E_z(a, \phi)\cos 2\nu\phi d\phi \quad \text{for } \nu = 1, 2 \ldots \quad . \quad (11b)$$

$$E_{z2}(a, \phi) = \sum_{\nu=0}^{\infty} B_{2\nu+1} \cos (2\nu + 1)\phi \quad . \quad (12)$$

where
$$B_{2\nu+1} = \frac{2}{\pi}\int_{-\pi/2}^{\pi/2} E_z(a, \phi)\cos (2\nu + 1)\phi d\phi \quad . \quad (13)$$

Using eqns. (6), (8), (10) and (12), and taking into account the facts that, for region 1, the field cannot be infinite at $r = 0$, and for region 2, E_z must be zero at $r = b$,

$$E_{z1}(r, \phi) = \sum_{\nu=0}^{\infty} A_{2\nu}\frac{J_{2\nu}(\beta r)}{J_{2\nu}(\beta a)} \cos 2\nu\phi \quad . \quad (14)$$

and
$$E_{z2}(r, \phi) = \sum_{\nu=0}^{\infty} B_{2\nu+1}\frac{Q_{2\nu+1}(\beta r, \beta b)}{Q_{2\nu+1}(\beta a, \beta b)} \cos (2\nu + 1)\phi \quad . \quad (15)$$

Using eqn. (7),

$$H_{\phi 1}(r, \phi) = -\frac{j\omega\epsilon}{\beta}\sum_{\nu=0}^{\infty} A_{2\nu}\frac{J'_{2\nu}(\beta r)}{J_{2\nu}(\beta a)} \cos 2\nu\phi \quad . \quad (16)$$

and

$$H_{\phi 2}(r, \phi) = -\frac{j\omega\epsilon}{\beta}\sum_{\nu=0}^{\infty} B_{2\nu+1}\frac{Q'_{2\nu+1}(\beta r, \beta b)}{Q_{2\nu+1}(\beta a, \beta b)} \cos (2\nu + 1)\phi \quad . \quad (17)$$

Again, as in TM_{n10} modes there is no variation of field along the z-axis; equating the surface integral of the Poynting vector across the dividing surface and using the conditions of symmetry,

$$\int_{-\pi/2}^{\pi/2} E_{z1}(a, \phi)H_{\phi 1}(a, \phi)d\phi = \int_{-\pi/2}^{\pi/2} E_{z2}(a, \phi)H_{\phi 2}(a, \phi)d\phi \quad . \quad (18)$$

Substituting for $E_{z1,2}(a, \phi)$ and $H_{\phi 1,2}(a, \phi)$ from eqns. (10), (12), (16) and (17) into eqn. (18); substituting for $A_{2\nu}$ and $B_{2\nu+1}$ from eqns. (11) and (13); integrating term-wise, and cancelling out the factor $-j\omega\epsilon/\pi\beta$ from all the terms, one obtains:

$$\frac{1}{2}\left[\int_{-\pi/2}^{\pi/2} E_z(a, \phi)d\phi\right]^2 \frac{J'_0(\beta a)}{J_0(\beta a)} + \sum_{\nu=1}^{\infty}\left[\int_{-\pi/2}^{\pi/2} E_z(a, \phi)\cos 2\nu\phi d\phi\right]^2 \frac{J'_{2\nu}(\beta a)}{J_{2\nu}(\beta a)} = \sum_{\nu=0}^{\infty}\left[\int_{-\pi/2}^{\pi/2} E_z(a, \phi)\cos (2\nu + 1)\phi d\phi\right]^2 \frac{Q'_{2\nu+1}(\beta a, \beta b)}{Q_{2\nu+1}(\beta a, \beta b)} \quad . \quad (19)$$

Transposing all terms except the first one to the right-hand side, and dividing by $\frac{1}{2}\left[\int_{-\pi/2}^{\pi/2} E_z(a, \phi)d\phi\right]^2$, eqn. (1) is obtained.

(9.2) Variational Expression

Substituting the definitions of Y^0, ΔY, $E_z^0(a, \phi)$ and $\Delta E_z(a, \phi)$ into eqn. (19), and omitting the explicit writing of (a, ϕ),

$$(Y^0 + \Delta Y)\left[\int_{-\pi/2}^{\pi/2}(E_z^0 + \Delta E_z)d\phi\right]^2 = \sum_{\nu=0}^{\infty}\left[\int_{-\pi/2}^{\pi/2}(E_z^0 + \Delta E_z)\cos(2\nu+1)\phi d\phi\right]^2 \frac{Q'_{2\nu+1}(\beta a, \beta b)}{Q_{2\nu+1}(\beta a, \beta b)} - \sum_{\nu=1}^{\infty}\left[\int_{-\pi/2}^{\pi/2}(E_z^0 + \Delta E_z)\cos 2\nu\phi d\phi\right]^2 \frac{J'_{2\nu}(\beta a)}{J_{2\nu}(\beta a)} \quad . \quad . \quad . \quad (20)$$

Expanding the squares, integrating term-wise, and using the definition of Y^0, one can cancel out all terms involving $\left(\int_{-\pi/2}^{\pi/2} E_z^0 \cos 2\nu\phi d\phi\right)^2$ and $\left[\int_{-\pi/2}^{\pi/2} E_z^0 \cos(2\nu+1)\phi d\phi\right]^2$.

Again, substituting the condition $H_{\phi 1}(a, \phi) = H_{\phi 2}(a, \phi)$, substituting for $A_{2\nu}$ and $B_{2\nu+1}$ from eqns. (11) and (13) and substituting Y^0 for $J_0'(\beta a)/J_0(\beta a)$ [given by eqn. (1), for the case where β has the correct value], one can cancel out from the expansion for eqn. (20) all terms involving

$$\left(\int_{-\pi/2}^{\pi/2} E_z^0 \cos 2\nu\phi d\phi\right)\left(\int_{-\pi/2}^{\pi/2} \Delta E_z \cos 2\nu\phi d\phi\right)$$

and $\left[\int_{-\pi/2}^{\pi/2} E_z^0 \cos(2\nu+1)\phi d\phi\right]\left[\int_{-\pi/2}^{\pi/2} \Delta E_z \cos(2\nu+1)\phi d\phi\right]$

Thus one is left with

$$\Delta Y\left(\int_{-\pi/2}^{\pi/2} E_z d\phi\right)^2 + Y^0\left(\int_{-\pi/2}^{\pi/2} \Delta E_z d\phi\right)^2 = \sum_{\nu=0}^{\infty}\left[\int_{-\pi/2}^{\pi/2} \Delta E_z \cos(2\nu+1)\phi d\phi\right]^2 \frac{Q'_{2\nu+1}(\beta a, \beta b)}{Q_{2\nu+1}(\beta a, \beta b)} - \sum_{\nu=1}^{\infty}\left(\int_{-\pi/2}^{\pi/2} \Delta E_z \cos 2\nu\phi d\phi\right)^2 \frac{J'_{2\nu}(\beta a)}{J_{2\nu}(\beta a)} \quad . \quad . \quad . \quad . \quad (21)$$

As Y is independent of the scale of E, the constant term in the trial field may be considered to be of the correct magnitude, the other terms being adjusted relative to it. Thus without loss of generality one can put

$$\int_{-\pi/2}^{\pi/2} \Delta E_z d\phi = 0$$

Dividing by $\left(\int_{-\pi/2}^{\pi/2} E_z d\phi\right)^2$, eqn. (2) is obtained.

621.314.58 : 621.382.233

The Institution of Electrical Engineers
Paper No. 3642 E
Sept. 1961

A DESIGN BASIS FOR SILICON-CONTROLLED RECTIFIER PARALLEL INVERTERS

By R. H. MURPHY, B.Sc.(Eng.), Graduate, and K. P. P. NAMBIAR, M.Sc., Associate Member.

(*The paper was first received 2nd February, and in revised form 27th April,* 1961.)

SUMMARY

The operation of the silicon-controlled rectifier in the basic form of parallel inverter circuits with resistive loads is investigated, and analysis is carried out to determine the conditions for optimum performance. Two fundamental parameters in terms of time-constants are introduced, one relating to the ballast inductance and the other to the commutating capacitance, as opposed to the method of defining design parameters in terms of a required triggering period hitherto employed in analysing thyratron circuits. The analysis shows that the basic circuit has three main modes of operation, two of which are undesirable. Within the range of the desirable mode square-wave and sine-wave operations are found to be feasible, and a procedure is given for determining the optimum triggering points in terms of time-constants. Expressions are derived for the maximum repetition rates for the generation of both types of waveforms in terms of relevant time-constants and turn-off times of the silicon-controlled rectifiers; with currently available high-current silicon-controlled rectifiers repetition rates of the order 6kc/s for square-wave operation and 12kc/s for sine-wave operation are found to be readily obtainable. With low-current diffused types, the corresponding repetition rates would be of the order of 35 and 70kc/s, respectively. The complete design procedure is illustrated by applying the analytical results to the practical design of a d.c./d.c. convertor, which is essentially an inverter followed by a bridge-rectifier filter unit.

(1) INTRODUCTION

In applications such as aircraft, missile and control instrumentation, power-transistor push-pull inverters have effectively replaced rotary inverters, vibrator-transformer inverters and other mechanical methods of converting d.c. battery supplies to higher-voltage a.c. supplies. It now seems almost certain that a relatively new device, the silicon-controlled rectifier (s.c.r.) will replace the transistor in this field and extend the range of these low-power inverters into the kilowatt region.

The s.c.r. is the semiconductor analogue of the thyratron. It is a 3-terminal, 4-layer *p–n–p–n* device with the property of being able to switch currents of the order of several tens of amperes upon application of a current pulse of a few milliamperes to its gate. It is normally turned off by external reduction of the forward current, e.g. by momentary reversal of the cathode-anode potential.

Apart from the more obvious advantages such as higher current rating, size and weight reduction, and robustness, the s.c.r. possesses two distinct advantages over the thyratron in this application. First, the turn-off time of the device is very much less than the deionization time of the thyratron, so that efficient operation of inverters may be achieved at very high repetition rates. This leads to size and weight reduction in the transformer and in the smoothing components if the inverter is followed by a rectifier-filter unit for d.c. reconversion. Secondly, the forward voltage drop when conducting is of the order of 1 V, compared with approximately 10 V of the thyratron, and hence the unit is potentially more efficient. This relatively low voltage drop makes the device very well suited for conversion of very low direct voltages to higher alternating and direct voltages. It possesses the disadvantage that the forward and reverse breakdown voltages are relatively lower, so that precautions should be taken to protect the device from excessive switching transients.

The first basic analysis of thyratron parallel-inverter circuits was carried out by Wagner.[1] In the following Sections the operation of s.c.r. circuits is analysed by adopting a very similar approach to that of Wagner but at the same time simplifying the procedure by means of Laplace transformation technique.[2] In addition, the analysis is extended to specific applications taking into consideration the appropriate characteristics of the s.c.r. An example of an inverter design for use in a d.c./d.c. convertor, one of the many potential applications of the s.c.r., is then given as an illustration of the design procedure.

(2) BASIC CIRCUIT OPERATION

The basic circuit is shown in Fig. 1. Positive-current trigger pulses are applied alternatively to the s.c.r.s at a rate corre-

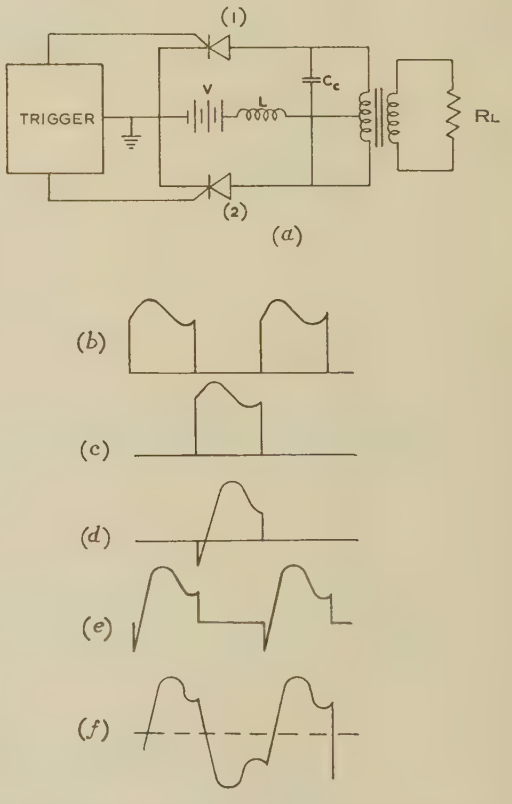


Fig. 1.—Basic inverter circuit and associated waveforms.
(*a*) Schematic of inverter.
(*b*) Current in s.c.r. 1.
(*c*) Current in s.c.r. 2.
(*d*) Voltage across s.c.r. 1.
(*e*) Voltage across s.c.r. 2.
(*f*) Output waveform.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Mr. Murphy and Mr. Nambiar are with Transitron Electronic, Ltd.

sponding to the desired frequency of operation. Briefly, the operation is as follows:

Assuming that s.c.r. 1 is conducting and has just reached steady-state conditions, its anode will be approximately at earth potential, the forward voltage during conduction being of the order of 1 V. The centre tap of the transformer will be at $+V$ above earth, and the anode of s.c.r. 2 instantaneously at $+2V$ by transformer action.

If s.c.r. 2 is now triggered 'on', the commutating capacitor C_c will try to discharge via the low-resistance path provided by the s.c.r.s, i.e. in the forward direction through s.c.r. 2 and the reverse direction through s.c.r. 1. However, since s.c.r. 1 cannot pass more than an instantaneous pulse of reverse current sufficient for carriers to diffuse out of its end junctions, it will be turned off and assume a blocking condition.

In the process of turning on, the anode of s.c.r. 2 falls to approximately earth potential and the capacitor, in effect, applies a voltage of $-2V$ from a low-impedance source to the anode of s.c.r. 1, this being an extremely fast way of turning it off. Thus when its anode swings positive again, s.c.r. 1 will remain 'off', and after the switching transient has died out, will try to stabilize with its anode at $+2V$ as in the case of s.c.r. 2 in the preceding operation. It is then ready for triggering in order that the half-cycle may repeat with s.c.r. 1 conducting.

The waveforms which result from this mode of operation are as sketched in Fig. 1(*b*)–(*f*). There are actually two other possible modes of operation; the following analysis is to show how these may arise, why they are undesirable, and what design considerations need to be taken into account in order to ensure that the inverter operates in the correct mode described.

(3) CIRCUIT ANALYSIS

Since only one s.c.r. conducts at a given time the derivation of the equivalent circuit for the inverter will be as shown in Fig. 2. Assuming that the transformer ratio is $1 + 1 : n$,

$$C = 4C_c \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

$$R = \frac{R_L}{n^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Also

$$\text{Voltage across } C_c = 2 \times (\text{Voltage across C}) \quad . \quad . \quad (3)$$

In the desired mode of operation, the circuit conditions existing when one s.c.r. is about to be triggered (the other s.c.r. having been conducting for the previous half-cycle) are as follows:

(*a*) A current of magnitude V/R is flowing in the ballast inductor.

(*b*) The commutating capacitor C_c is charged to a voltage of $+2V$. Because of transformer action, the equivalent capacitor is charged to a voltage V [see eqn. (3)].

These conditions are evident from the description of operation of the circuit and will be further justified by the analysis to follow, which defines the necessity for the boundary conditions to match at the start and finish of a half-cycle. Applying these initial conditions, the equivalent operational circuit at the start of a half-cycle may be evolved as shown in Fig. 2(*c*).

(3.1) Variation of Current in S.C.R. during Half-Cycle

From Fig. 2(*c*),

$$i(p) = \left(\frac{V}{p} + \frac{CR}{1+pCR} \cdot V + \frac{LV}{R}\right)\left(\frac{1+pCR}{p^2LCR + pL + R}\right) \quad . \quad (4)$$

Putting

$$CR = T_c \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

$$\frac{L}{R} = T_L \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

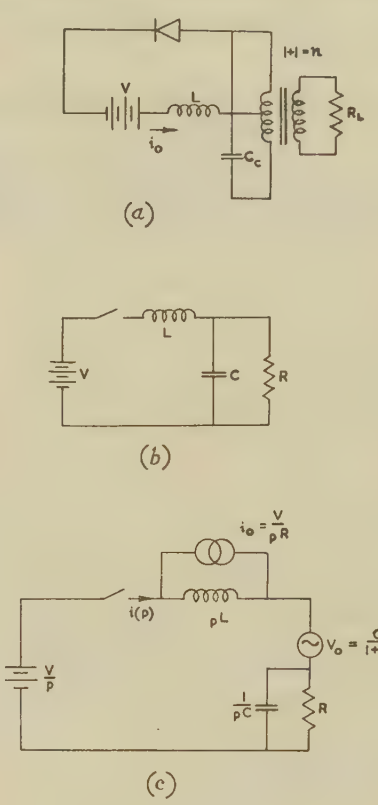


Fig. 2.—Equivalent circuit of the inverter.

(*a*) Circuit presented to one s.c.r.
(*b*) Equivalent of (*a*) assuming s.c.r. to be a perfect switch.
(*c*) Operational equivalent of (*b*) showing initial conditions when s.c.r. is about to be triggered.

Eqn. (4) reduces to

$$i(p) = \frac{V}{R} \frac{\left[p^2 + p\left(\frac{1}{T_c} + \frac{2}{T_L}\right) + \frac{1}{T_cT_L}\right]}{p\left(p^2 + \frac{p}{T_c} + \frac{1}{T_cT_L}\right)} \quad . \quad . \quad (7)$$

The current through the s.c.r. at any instant during the half-cycle may be determined by evaluating the inverse transform of eqn. (7), and is given by

$$i(t) = \frac{V}{R} \times$$

$$\left[1 + \frac{2\frac{T_c}{T_L}}{\left(\frac{T_c}{T_L} - \frac{1}{4}\right)^{1/2}} \exp\left(-\frac{t}{2T_c}\right) \sin\left(\frac{1}{T_c}\right)\left(\frac{T_c}{T_L} - \frac{1}{4}\right)^{\frac{1}{2}} t\right]$$

. . . . . (8)

This may be simplified by introducing a dimensionless factor Q such that

$$Q = \left(\frac{T_c}{T_L} - \frac{1}{4}\right)^{1/2} \quad . \quad . \quad . \quad . \quad . \quad (9)$$

Eqn. (8) then becomes

$$i(t) = \frac{V}{R}\left[1 + \frac{2(Q^2 + \frac{1}{4})}{Q} \exp\frac{-t}{2T_c} \sin\frac{Q}{T_c}t\right] \quad . \quad (10)$$

For all values of $T_c/T_L > \frac{1}{4}$, Q is a real constant, and the current waveform $i(t)$ from eqn. (10) may be represented by a

damped sine-wave oscillation about the reference level $i_{(t=0)} = V/R$ as illustrated in Fig. 3.

(3.2) Condition for Maintaining Conduction throughout the Half-Cycle

Since an s.c.r. cannot pass appreciable current in the reverse direction it will be cut off if the oscillations are so great that the curve in Fig. 3 cuts the $i(t) = 0$ axis at any point. The limitation

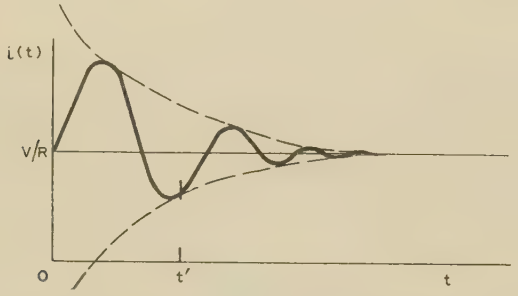


Fig. 3.—Current flow through s.c.r. during conduction.

Exponential decay is given by $\frac{2(Q^2+\frac{1}{4})}{Q}\varepsilon^{-t/2T_c}$ and the angular frequency by Q/T_c.

is evidently on the point at which the first minimum occurs, and the condition for this point just to touch the axis will now be found.

From eqn. (10) the value of t' at which the first minimum will occur is given approximately by

$$\frac{Q}{T_c}t' = \frac{3\pi}{2}, \quad \text{i.e.} \quad \frac{t'}{2T_c} = \frac{3\pi}{4Q} \quad . \quad . \quad . \quad . \quad (11)$$

Substituting eqn. (11) in eqn. (10) gives

$$i(t') = \frac{V}{R}\left[1 - 2\left(\frac{Q^2+\frac{1}{4}}{Q}\right)\exp\left(-\frac{3\pi}{4Q}\right)\right]$$

Hence $i(t') = 0$ when

$$2\left(\frac{Q^2+\frac{1}{4}}{Q}\right)\exp\left(\frac{-3\pi}{4Q}\right) = 1$$

i.e. when

$$\log_\varepsilon \frac{2(Q^2+\frac{1}{4})}{Q} = \frac{3\pi}{4Q} \quad . \quad . \quad . \quad . \quad . \quad (12)$$

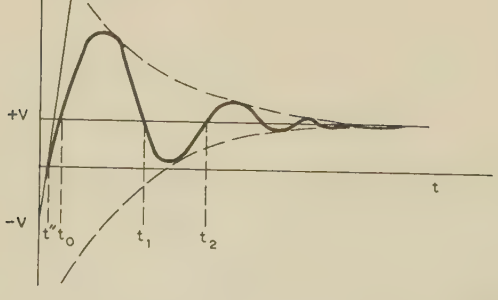


Fig. 4.—Variation of commutating voltage during the half-cycle.

Eqn. (12) may be solved graphically by assuming different values for Q and substituting these in the left- and right-hand expressions. A curve for each expression plotted against $(Q^2 + \frac{1}{4})$ is shown in Fig. 5. The limiting condition occurs

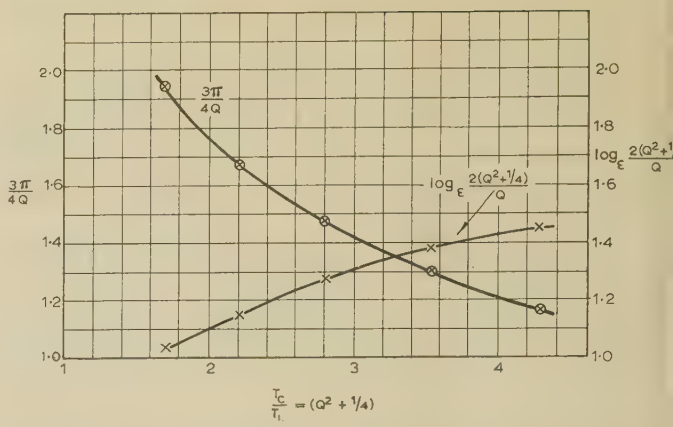


Fig. 5.—Graphical solution for the limiting value of T_c/T_L.

at the point of intersection of the two curves, and is shown to correspond approximately to 3·28 on the $(Q^2 + \frac{1}{4})$ axis. Thus for continuous operation of the s.c.r.,

$$(Q^2+\tfrac{1}{4}) = \frac{T_c}{T_L} < 3{\cdot}28 \quad . \quad . \quad . \quad . \quad (13$$

For the purpose of approximation, the slight influence of the exponential decay on the position of the first minimum has been ignored.

The method adopted by Wagner involves determining the point of the first maximum and adding π to this period (on an angular-frequency basis) to evaluate the first minimum; substitution for this value and solution by trial and error yield

$$\frac{T_c}{T_L} < 3{\cdot}24 \quad . \quad . \quad . \quad . \quad . \quad . \quad (14$$

Assuming this to be a slightly more accurate value, the conditions for continuous conduction with the type of solution illustrated in Fig. 3 becomes

$$0{\cdot}25 < \frac{T_c}{T_L} < 3{\cdot}24 \quad . \quad . \quad . \quad . \quad . \quad (1$$

(3.3) Variations of Commutating Voltage during Half-Cycle

From Fig. 2(*c*) the operational form of the voltage across the load is given by

$$v(p) = VR\frac{\left(\frac{1}{p} + \frac{CR}{1+pCR} + \frac{L}{R}\right)}{(p^2LCR + pL + R)} \quad . \quad . \quad . \quad (16$$

This reduces to

$$v(p) = V\left[\frac{1}{p(1+pT_c)} + \frac{2T_c}{(1+pT_c)p^2T_cT_L + pT_L + 1}\right] \quad . \quad (1$$

The voltage across the load at any instant during the half-cycle may be evaluated by the inverse transform of eqn. (17) and is given by

$$v(t) = V\left[\exp\left(\frac{-t}{2T_c}\right)\left(-2\cos\frac{Qt}{T_c} + \frac{1}{Q}\sin\frac{Qt}{T_c}\right) + 1\right] \quad . \quad (1$$

where the dimensionless factor Q defined by eqn. (9) is introduced as before. Provided that T_c/T_L satisfies the conditions expressed by eqn. (15), $v(t)$ is of the form shown in Fig. 4, this being a graphical representation of eqn. (18).

(3.4) Condition for the Conducting S.C.R. to Turn Off

The s.c.r. must be reverse-biased long enough for it to be able to turn off completely. Hence, from Fig. 4 the limiting condition is that $t'' > t_{off}$, where t_{off} is the maximum turn-off time specified in the data sheets for the particular s.c.r.s. This condition is most easily found by ignoring the curvature of the waveform between $t = 0$ and $e = 0$; determining the gradient of the curve of eqn. (18) at $t = 0$,

$$\frac{dv}{dt_{(t=0)}} = \frac{2V}{T_c} \quad . \quad . \quad . \quad . \quad . \quad (19)$$

Hence, from Fig. 4 and eqn. (18):

$$\frac{V}{t''} = \frac{2V}{T_c}, \text{ i.e. } t'' = \frac{T_c}{2}$$

i.e. the s.c.r. will turn off successfully provided that

$$\frac{T_c}{2} > t_{off} \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

Since the effect of the curvature of the first part of the curve has been neglected in the analysis, the condition given by eqn. (19) allows sufficient margin for the completion of the turn-off process; this is expressly desirable since inability to turn off would lead to a short-circuit condition.

(3.5) Square-Wave Operation

The boundary conditions at the beginning and end of a half-cycle in Figs. 3 and 4 only match exactly if the switching transient has died out before triggering occurs. If this is so, the inverter operates with approximately a square-wave output, and this is an efficient* method of operation because commutation losses occupy only a small fraction of the half-cycle. There is, however, a maximum frequency of operation, limited by the need for the exponential to die out before the end of the half-cycle and yet turn the s.c.r. off successfully, i.e. satisfy eqn. (20).

The retriggering may be effectively applied when the exponential has decayed to less than 5% of its initial value, i.e. within 5% of the reference-voltage level of V or current level of V/R, so that from eqns. (10) and (17) we have

$$\exp\left(-\frac{t}{2T_c}\right) = 0{\cdot}05 = \exp(-3)$$

The half-cycle period T is therefore given by

$$T > 6T_c \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

or the maximum repetition rate f_{max} is given by

$$f_{max} = \frac{1}{2T_{min}} = \frac{1}{12CR} \quad . \quad . \quad . \quad . \quad . \quad (22)$$

(3.6) Sine-Wave Operation

Boundary conditions approximately match, and therefore operation is possible if triggering occurs at the points t_1, t_2, etc., on Figs. 3 and 4. The point of main interest is t_1, for this corresponds to the highest possible frequency of operation and gives approximately a sine-wave output voltage.

Thus if $t_n = t_0, t_1, t_2$, etc., putting $v(t = t_n) = +V$ in eqn. (18)

$$\frac{1}{Q}\sin Q\frac{t_n}{T_c} = 2\cos Q\frac{t_n}{T_c}$$

i.e.

$$\frac{t_n}{T_c} = \frac{1}{Q}(\tan^{-1} 2Q + n\pi)$$

* This assumes that the transformer handles the square wave perfectly and that a square wave is a suitable output so far as the circuit to be driven is concerned, e.g. in a converter the method would not only be efficient but would produce less ripple.

where n is an integer as illustrated in Fig. 4. So that

$$\frac{t_1}{T_c} = \frac{1}{Q}(\tan^{-1} 2Q + \pi) \quad . \quad . \quad . \quad . \quad (23)$$

$\frac{t_1}{T_c}$ is therefore a function of Q only.

A graph of this function against $T_c/T_L(= Q^2 + \frac{1}{4})$ in the range expressed by eqn. (15) is given in Fig. 6, and the method

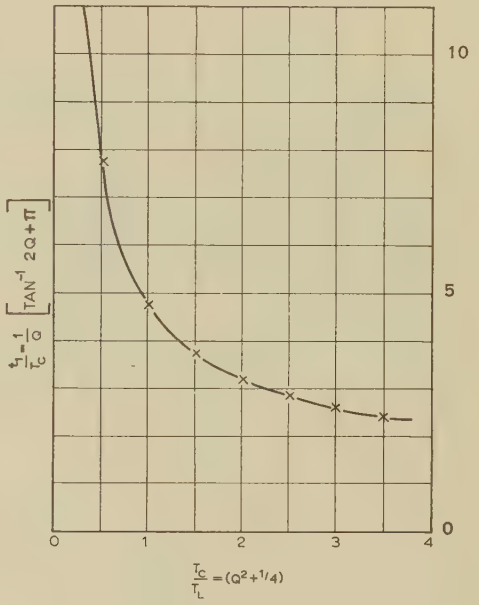


Fig. 6.—Graph of t_1/T_c versus T_c/T_L for determining parameters for approximate sine-wave operation.

of determining optimum values of T_c and T_L from the curve for sine-wave operation at a particular frequency is included in Section 5.

The upper limit in eqn. (15) is no longer strictly applicable since triggering occurs before the first minimum on Fig. 3. In practice the limiting value for T_c/T_L is found to be rather less than 3·24, since, as discussed in Section 4, starting becomes more and more unreliable as the ratio approaches this value.

(4) STARTING CONDITIONS

If both s.c.r.s are initially 'off', and then one is triggered, the initial conditions in Fig. 2(*c*) do not apply; in fact, the initial conditions are quiescent.

Thus, equating the two voltage terms corresponding to the initial condition in eqns. (4) and (16) to zero and conducting a similar analysis, we have

$$i(t) = \frac{V}{R}\left[1 + \frac{(Q^2 + \frac{1}{4})}{Q}\exp\frac{-t}{2T_c}\sin\left(\frac{Qt}{T_c} + \psi_1\right)\right] \quad . \quad (24)$$

where $\psi_1 = \tan^{-1} 2Q - \tan^{-1}(-2Q)$

$$v(t) = V\left[1 + T_c\frac{(T_cT_L)^{\frac{1}{2}}}{Q}\exp\frac{-t}{2T_c}\sin\left(\frac{Qt}{T_c} + \psi_2\right)\right] \quad . \quad (25)$$

where $\psi_2 = \tan^{-1}(-2Q)$.

From the above equations it will be seen that the boundary condition for starting is mismatched at the end of a half-cycle as a result of the phase difference between the two expressions. For optimum starting condition, the phase difference $(\psi_1 - \psi_2)$ should be zero, or alternatively the switching transient repre-

sented by the exponential terms should have decayed approximately to zero [see eqn. (20)]. Sine-wave mode of operation is thus inherently difficult to start.

The phase difference does approach zero on the fringe of the desired operating range as Q tends to zero, i.e. $T_c/T_L \rightarrow 0{\cdot}25$.

As this condition is approached, however, the quality of the sine wave deteriorates, and eventually at $T_c/T_L = 0{\cdot}25$ the waveform consists of an increasing exponential followed by an abrupt trail.

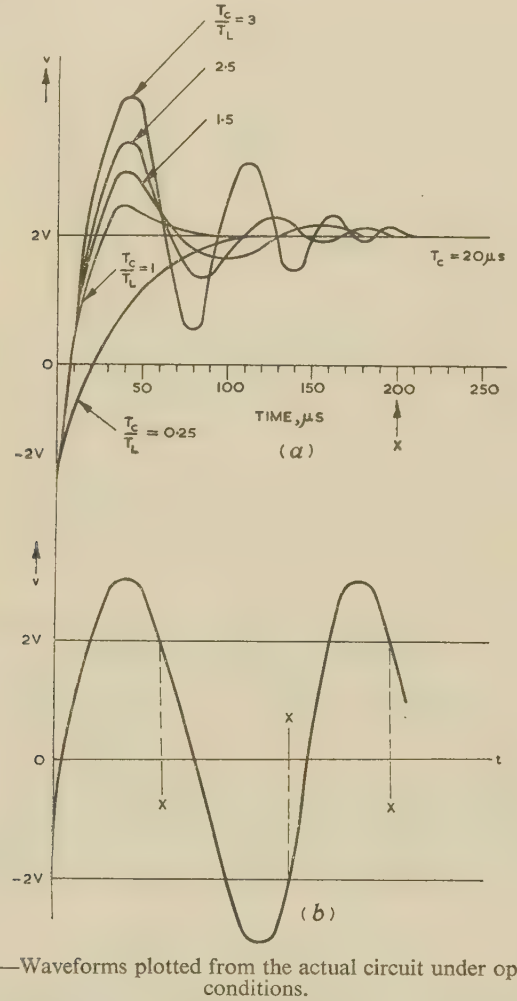


Fig. 7.—Waveforms plotted from the actual circuit under operating conditions.

(*a*) Oscillograms of the commutating voltage for various values of T_c/T_L with $T_c = 20\,\mu s$.
(*b*) Generation of approximate sine wave by triggering before the decay of the exponential is completed ($T_c/T_L = 2{\cdot}5$).
X. Trigger.

Fig. 7(*a*) shows oscillograms of the various voltage waveforms for different values of T_c/T_L with T_c constant, and thus is an aid in finding compromise values for these conflicting requirements. An example of such a compromise is illustrated in Fig. 7(*b*).

(5) SUMMARY OF ANALYTICAL RESULTS

(*a*) The desired mode of operation occurs when

$$0{\cdot}25 < \frac{T_c}{T_L} < 3{\cdot}24 \text{ [see eqn. (15)]}$$

If $T_c/T_L < 0{\cdot}25$ the output waveform has an exponential leading edge, the switching transient being heavily damped [see Fig. 7(*a*)]. This also results in a decrease in efficiency and inability of the inverter to adjust itself to abrupt changes of load because the relatively large ballast inductor entailed produces large transients across the s.c.r.s which are then liable to suffer reverse failure. If $T_c/T_L > 3{\cdot}24$ the current flows in pulses from the supply, and hence the mode is inefficient and generally undesirable.

(*b*) For square-wave operation, which is for most purposes the most efficient and desirable mode, the frequency of operation is limited by the fact that the half-cycle period T must be greater than approximately $6T_c$ [eqn. (21)], where T_c itself must be greater than $2t_{off}$ [eqn. (20)]. For example, t_{off} with present-day 20 A s.c.r.s is in the region of $7\,\mu s$. Hence the maximum frequency for square-wave operation using this type of s.c.r. is approximately* 6 kc/s.

It should be noted that operating at a high frequency allows size and weight reductions in the transformer (and smoothing components if the unit is part of a convertor), but these are offset by greater commutation losses and hence lower efficiency and poorer regulation. Probably the most practical operating frequency for this type of inverter is in the region of 1–3 kc/s, where efficiencies of well over 90% may be obtained with relatively small components.

(*c*) For approximate sine-wave operation T_c/T_L should be low for reliability and high for quality and efficiency. If suitable starting procedures[3] are adopted, reliable operation up to $T_c/T_L = 2{\cdot}5$ may be obtained. The waveform is then as shown in Fig. 7(*b*). For operation in this region, other design specifications are as follows:

From Fig. 6, when $T_c/T_L = 2{\cdot}5$, $t_1/T_c = 2{\cdot}94$

if the desired frequency is f cycles per second $t_1 = \frac{1}{2f}$, so that

$$T_c = \frac{1}{5{\cdot}88f}$$

For example, at $f = 2{\cdot}5$ kc/s,

$$T_c = 68\,\mu s$$
$$T_L = 27\,\mu s$$

Here, T_c easily satisfies the condition $T_c > 2t_{off}$ for the 20 A s.c.r.s already quoted. For the highest possible frequency of sine-wave operation with this compromise value for T_c/T_L and the device turn-off time of $7\,\mu s$ (i.e. $T_c = 14\,\mu s$) $\frac{1}{2f_{max}T_c} = 2{\cdot}94$ so that $f_{max} = 12$ kc/s (compared with $f_{max} = 6$ kc/s for square-wave operation).

(6) EXAMPLE OF SPECIFIC DESIGN

The design of an inverter, typical of the immediate potential of available s.c.r.s, will now be discussed in detail. In practical operation, this inverter is followed by a bridge rectifier-filter unit, and the whole assembly thus becomes a convertor. The rectifier-filter design conforms with conventional techniques and will not be detailed. However, one point worth noting is that the most suitable filter is a choke-input type, for the following reasons:

(*a*) This type of filter takes a continuous current over the whole half-cycle and thus will not load the inverter on a pulse basis as would a capacitor input filter.
(*b*) A choke-input filter is the more efficient.

The effect of the reflected inductance on the inverter operation is to neutralize some of the commutating capacitance, which

* Employing low-current diffused s.c.r.s with a turn-off time of the order of 1·2μs, repetition rates as high as 35 kc/s may be obtained.

should therefore be increased slightly unless the turns ratio of the transformer is so large as to make this unnecessary.

Assume a convertor is required to operate from a 27·5 V battery and provide a 250 V 125 mA output and a 170 V 125 mA output. For reasons discussed in Section 5, 2·5 kc/s is considered a suitable operating frequency.

The specification for the transformer is illustrated in Fig. 8. The transformer design is quite critical because residual magnetization can cause random starting difficulties. The transformer is designed to operate at only about one-third of the saturation flux density specified for the core. Bearing this in mind, the usual design techniques may be adopted, and the result of this particular design is detailed in Fig. 8.

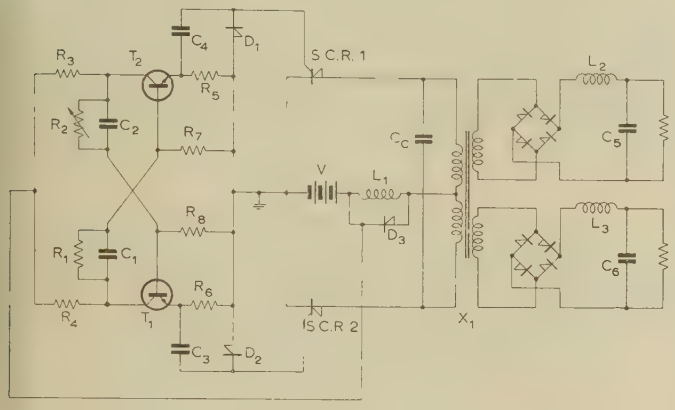


Fig. 8.—Circuit diagram of a d.c./d.c. convertor.

Resistors.
R_1: 4.7 kΩ
R_2: 5 kΩ (variable)
R_3, R_4: 330 Ω
R_5, R_6: 220 Ω
R_7, R_8: 2·2 kΩ

Inductors.
L_1: 120 μH
L_2, L_3: 350 mH

Capacitors.
C_c: 0·75 μF
C_1, C_2: 0·2 μF
C_3, C_4: 0·1 μF
C_5, C_6: 1 μF

Semiconductor devices.
D_1, D_2: T8G S.C.R. 1 and S.C.R. 2: TCR 1005
D_3: 1N202
Diode bridge: Four TR401 each.

Transformer T_1.
Core: Super Stalloy 0·014 in.
Laminations: ¾ in No-waste E and I type 70.
Primary: bifilar-wound 28 turns each half of No. 18 s.w.g. enamelled copper.
First secondary: 306 turns No. 33 s.w.g. enamelled copper.
Second secondary: 208 turns No. 33 s.w.g. enamelled copper.

Assuming 85% efficiency, the primary current taken from the supply is $\frac{75\,\text{V A}}{0{\cdot}85 \times 27{\cdot}5\,\text{V}} = 3{\cdot}2\,\text{A}$ [see Fig. 6]. Therefore, the s.c.r.s are required to handle average currents of 1·6 A, and withstand peak reverse voltages of 55 V (=2 V) in addition to the switching transients. Thus 100 V 5 A types will be suitable.

Assume that, for these, t_{off} (typical) = 7 μs. About 50% should be added to allow for production spreads and to make the design less critical. Thus $t_{off} = 10\,\mu s$. 2·5 kc/s corresponds to a half-cycle period of 200 μs.

Hence $T \geqslant 6T_c$ [from eqn. (21)] gives $T_c < 33\,\mu s$

$T_c > 2t_{off}$ [from eqn. (20)] gives $T_c > 20\,\mu s$

The load R reflected into the primary is R_{L1}/n_1^2 in parallel with R_{L2}/n_2^2. From the transformer details of Fig. 8, $R = 10\,\Omega$. $T_c = CR = 4C_cR$ must lie between 20 and 33 μs, i.e. C_c must lie between 0·5 and 0·825 μF. Taking a value towards the high end to allow for reflected inductance this corresponds to $T_c = 30\,\mu s$, $C_c = 0{\cdot}75\,\mu F$. As long as T_c/T_L lies between 0·25 and 3·24 the choice depends only upon the waveform desired (Fig. 5). In this case a value of 2·5 is chosen as a compromise between low ripple and high efficiency:

$$\frac{T_c}{T_L} = 2{\cdot}5$$

$$T_L = \frac{30}{2{\cdot}5} = 12\,\mu s$$

$$\frac{L}{R} = 12\,\mu s, \text{ i.e. } L = 120\,\mu H$$

(6.1) Trigger Circuit

Of the various methods available for triggering s.c.r.s the most suitable for this application is a simple 2-transistor astable multivibrator. A circuit designed by entirely conventional techniques is shown in the complete circuit diagram (Fig. 8). This provides a 2·5 kc/s square wave from a low-impedance source, which, on differentiation by the coupling-capacitor-s.c.r. input-impedance network produces positive pulses of about 10 V, 180° out of phase on the gates of the s.c.r.; the negative pulses produced on differentiations are removed by diodes connected across the gate cathodes of the s.c.r. From the data sheets of these particular s.c.r.s it is seen that these pulses are suitable for triggering at all ambient temperatures within their rating.

(6.2) Transient Suppression

Relatively large values of the ballast inductor lead to large transients in the reverse direction across the s.c.r.s if the load is suddenly changed. A suitable diode may therefore be connected across this inductor as shown, to prevent these transients from exceeding the reverse breakdown voltage of the s.c.r.s (100 V) and hence damaging them.

(7) CONCLUSIONS

The analysis has shown that the basic inverter circuit has three main modes of operation, two of which are undesirable for reasons discussed in Section 5. Although the equivalent circuit evolved is very similar to that previously employed for thyratron inverters, the application of Laplace transformation technique has resulted in the derivation of an operational circuit which has not only simplified the analysis but introduced a significant extension to suit the particular requirements of s.c.r. inverters.

The main departure from the thyratron analysis due to Wagner has been the introduction of the time-constant parameters T_c and T_L, which are independent of the operating frequency, as opposed to defining design parameters in terms of a required triggering period. The approach has shown that two types of operation, square-wave and sine-wave, are feasible within the range of the desirable main mode and a procedure has been given for determining the optimum triggering points in terms of T_c and T_L. Only then has the process been reversed and the choice of T_c and T_L been made dependent on the operating frequency and turn-off period. This method enables the analysis to be carried out in a much more straightforward way; for example, the relevant frequencies of operation and the turn-off time determine T_c and the choice of T_L is dependent only on the choice of T_c/T_L appropriate to the waveform desirable from the point of view of efficiency, low ripple and purity of waveform.

It should be noted that this simplification has been made explicitly possible by the introduction of the s.c.r. to this type of circuit. In the thyratron circuit, since turn-off time is considerably longer, the method used for relating T_c to t_{off} [eqn. (20)] was inadmissible, because, far from being able to ignore the curvature of the first part of the waveform in Fig. 4, the circuit

had to be designed so that a pronounced curvature allowed the thyratron to turn off. In other words, since the deionization time of a thyratron can be anything up to 3 200 μs as compared with the order of 10 μs for the s.c.r., a damped switching-transient was necessary to ensure complete turn-off, and obtaining this type of waveform involved design based on both the inductance and capacitance parameters. The design, however, could only be carried out by the extremely tedious method of plotting current and voltage waveforms for various values of the frequency-dependent inductance and capacitance parameters and then plotting further curves of the required commutation ratio (corresponding to t'/T in this case) against the capacitance parameter for various fixed values of the inductance parameter; it is evident that a general analytical expression for this ratio derived from eqn. (18) would be so involved as to be of little value in practical circuit design.

The expressions derived for current and voltage and associated waveforms plotted in Figs. 3 and 4 form the design basis from which peak current and voltage levels, efficiency and ripple frequency may be readily evaluated.

The summary has shown how sine-wave operation parameters may be evaluated, for any frequency within the range determined by eqns. (15) and (20), by judicious use of Figs. 6 and 7. It should be noted that for reliable operation, however, starting procedures referred to in Reference 3 should be adhered to. The oscillograms plotted in Fig. 7 are useful as a rough guide to the type of waveform required either for sine- or square-wave operation in a circuit using the 20 A s.c.r. referred to, or an s.c.r. with any other rating provided that its turn-off time is of the same order (5–10 μs). However, by scaling the time axis accordingly these oscillograms may be modified to suit s.c.r.s with any turn-off time provided that the constant value of T_c chosen is twice the turn-off time, this being the optimum value of T_c from the point of view of efficiency.

The design of a typical inverter forming part of a d.c./d.c. convertor has been given and it is hoped that the incidental design considerations introduced and the results of the preceding analysis will form a useful practical design basis.

(8) ACKNOWLEDGMENTS

The authors wish to acknowledge their gratitude to Dr. David Bakalar for providing facilities for the circuit investigation and to the Directors of Transitron Electronic, Ltd., for permission to publish the paper.

Acknowledgment is also due to Mr. Roger Dellor of Transitron Applications Laboratory for his part in assisting the experimental investigations.

(9) REFERENCES

(1) WAGNER, C. F.: 'Parallel Inverter with Resistance Load', *Transactions of the American I.E.E.*, 1935, **54**, p. 1227.

(2) GOLDMAN, S.: 'Transformation Calculus and Electrical Transients' (Constable, 1949).

(3) Transitron Application Laboratory Technical Memo. No. 6. 'Designing SCR Inverters and Converters.'

621.382.233:621.374.3:621.398

The Institution of Electrical Engineers
Paper No. 3610 E
Sept. 1961

SOME CIRCUIT APPLICATIONS OF AVALANCHE DEVICES

By G. M. ETTINGER, Ph.D.

(*The paper was first received* 10*th January, and in revised form* 20*th March*, 1961.)

SUMMARY

Applications of *p–n–p–n* avalanche diodes in analogue computers, process control, oscillography and telemetery are discussed. These are based on a new voltage-controlled ramp-generator circuit of high linearity which employs *p–n–p–n* diodes and silicon transistors, and on a *p–n–p–n* variable-width pulse generator circuit. Some compatible transistor-*p–n–p–n* circuits are discussed which overcome the problem of the relatively high striking voltage of *p–n–p–n* diodes. It is pointed out that several of the above functions can be realized also with triode transistors operated in the avalanche mode, and details are given of a single-transistor sampling oscilloscope circuit operating up to nearly 2 Gc/s.

LIST OF SYMBOLS

α_1, α_2 = Collector efficiencies of centre junction in *p–n–p–n* diode for carriers injected at end junctions.
C = Timing capacitance, μF.
I_d = Maximum input current for avalanche circuit, A.
I_0 = Minimum input current to maintain repetitive operation in avalanche circuit, A.
I_H = Sustaining or hold current of *p–n–p–n* diode, A.
R = Series resistance for reduced-amplitude *p–n–p–n* ramp generator, Ω.
R_1, R_2 = Series resistances in variable-length *p–n–p–n* pulse generator, or leakage resistance, Ω.
R_s = Total current-limiting resistance in avalanche circuit, Ω.
T = Pulse length, sec.
t_d = Discharge time for avalanche circuit at current I_d, sec.
t_c = Charging time for avalanche circuit having series resistance R_s, sec.
τ = Time-constant for R_1, R_2 and C in parallel, sec.
v_1 = Instantaneous voltage across timing capacitor in variable-length *p–n–p–n* pulse generator, V.
V_c = Voltage across timing capacitor in variable-amplitude *p–n–p–n* ramp generator, V.
V_s = Breakdown voltage of *p–n–p–n* diode, V.

(1) GENERAL

(1.1) Operation of *p–n–p–n* Diodes

The operation of *p–n–p–n* diodes has been described by Shockley[1] and by Moll *et al.*[2] These devices have been employed for the generation of relaxation oscillations based on exponential capacitor charge and for the generation of pulses having millimicrosecond rise time. For the first application, the circuit simplicity obtainable with *p–n–p–n* avalanche diodes is important; the high switching speed, which cannot readily be equalled with conventional linear transistor circuits, is important for the second application.

Before proceeding to a consideration of *p–n–p–n* circuits, the characteristics of *p–n–p–n* diodes[3] will be briefly reviewed.

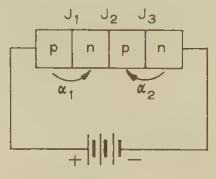


Fig. 1.—Basic *p–n–p–n* structure.

Consider the 4-layer structure shown in Fig. 1, where avalanche multiplication of minority carriers injected at the two end junctions can take place. The centre junction, initially reverse biased, has low leakage (typically 10^{-9} A for silicon at room temperature) as indicated in Fig. 2. With increasing supply voltage, the effective amplification $(\alpha_1 + \alpha_2)M$, where M is the avalanche multiplication factor, reaches unity, and avalanche breakdown takes place (B in Fig. 2). The centre junction

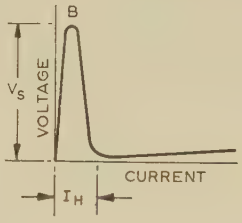


Fig. 2.—Current/voltage curve of *p–n–p–n* diode.

becomes forward-biased and the current through the device rises rapidly, limited only by the drop across the three forward junctions J_1, J_2 and J_3 and by any series resistance in the circuit. When the current is reduced below the hold current I_H, avalanche action is no longer possible and the *p–n–p–n* diode reverts to the high-resistance state. Both the breakdown voltage and the hold current can be controlled as device design parameters. The circuits described in the paper are based on the assumption of reasonable control of these parameters.

(1.2) Basic Circuit

Analogies are often drawn[3] between *p–n–p–n* ramp generators and gas-tube relaxation oscillators. The basic circuits are compared in Fig. 3. High frequency of operation is possible with

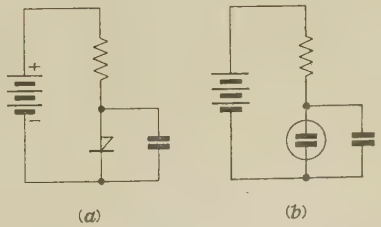


Fig. 3.—Relaxation oscillators.
(*a*) *p–n–p–n* relaxation oscillator.
(*b*) Gas-tube relaxation oscillator.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Dr. Ettinger was formerly with Marconi Instruments, Ltd., and is now with G. and E. Bradley, Ltd.

p–n–p–n diodes (up to about 500 kc/s repetitive) and avalanche transistors (up to about 10 Mc/s), compared with some tens of kilocycles per second only for gas tubes. Avalanche devices readily yield well-defined amplitudes of about 20–100 V.

Fig. 4 shows the waveforms obtainable from a simple *p–n–p–n*

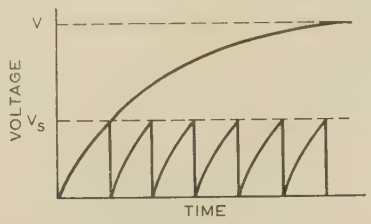


Fig. 4.—Waveform generated by *p–n–p–n* relaxation oscillator.

diode relaxation oscillator. To obtain reasonable linearity, the supply voltage must be considerably higher than the striking voltage. In order to achieve 1% and 2% linearity, for example, the supply voltage would have to be 50 and 25 times, respectively, the relaxation amplitude. The employment of such high supply voltages is often impracticable. A circuit arrangement described in the following Section overcomes this difficulty and is capable of yielding ultra-linear ramp functions, with the supply voltage only slightly higher than the striking voltage.

(2) LINEAR RAMP GENERATOR

Referring again to Fig. 3(*a*), it is apparent that the timing capacitor must be supplied with constant current in order to produce linear ramp functions. This is achieved in the circuit of Fig. 5 by the use of a common-base transistor with high

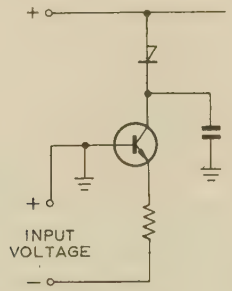


Fig. 5.—Transistor-controlled linear ramp generator (avalanche digitizer).

emitter resistance. This transistor discharges the capacitor at a constant rate until the voltage across the *p–n–p–n* diode reaches the breakdown value, when the capacitor is rapidly charged again from the positive supply voltage.

The current in the emitter and collector circuits of the transistor is proportional to the input voltage applied to the emitter resistance, subject to I_{c0}, V_{c0} and departure of α from unity. Hence the rate of change of voltage across the capacitor is substantially proportional to the input voltage. It follows, taking the amplitude of the function generated as constant, that the frequency should be proportional to the input voltage.* In practice, the linearity of the frequency/voltage curve can be maintained over input voltage ranges approaching 500 : 1. A typical frequency/voltage characteristic for a circuit based on Fig. 5 is shown in Fig. 6. Fig. 7 shows oscillograms obtained with the circuit of Fig. 5 as the input voltage was varied over the range 0·1–10 V.

* The effect of capacitor charging time on linearity is considered in Appendix 11.1.

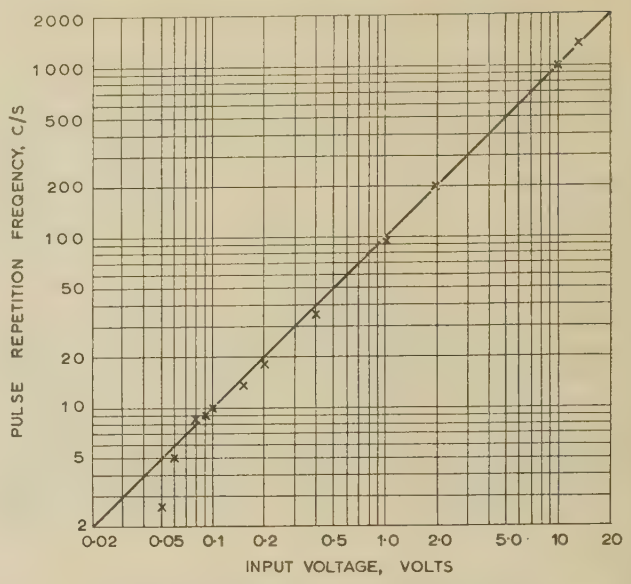


Fig. 6.—Voltage/frequency curve for avalanche digitizer based on circuit of Fig. 5.

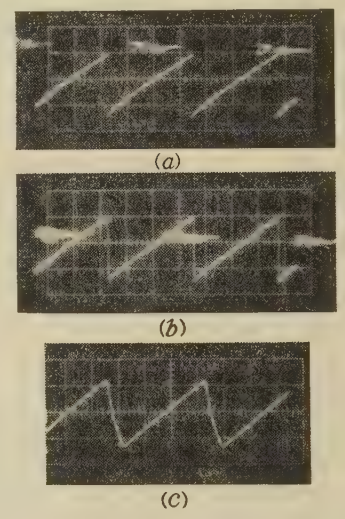


Fig. 7.—Output waveforms from transistor-controlled *p–n–p–n* generator (circuit of Fig. 5).

Vertical scale: 20 V./division.
(*a*) Input: 0·1 V. Horizontal scale: 1 ms/division.
(*b*) Input: 1 V. Horizontal scale: 100 μs/division.
(*c*) Input: 10 V. Horizontal scale: 10 μs/division.

The proportionality of frequency to input voltage for the circuit described leads to three groups of applications:

(*a*) Digitization, e.g. in telemetry, process control and other data-handling applications such as digital voltmeters.
(*b*) Generation of frequency modulation *per se*.
(*c*) Generation of fairly accurately defined time intervals, either on a repetitive basis or initiated by input trigger signals.

It is a corollary of the statement relating to linearity of the frequency/voltage characteristic over a 500 : 1 ratio that ramp functions of 0·1% linearity can be produced at maximum input voltage (see Appendix 11.2). This leads to applications in precision oscillography and in the generation of linear functions for analogue-computer and simulator work.

Temperature dependence of the breakdown voltage[4] for silicon *p–n–p–n* diodes is of the order of −0·02% per deg C. This lead-

to a proportional increase in frequency with temperature. Experimental digitizers, without temperature compensation, have shown a stability of approximately $\pm 0{\cdot}2\%$ of maximum frequency over the range 25–45° C.

(3) COMPATIBLE TRANSISTOR-*p–n–p–n* CIRCUITS

p–n–p–n diode circuits readily produce amplitudes of up to 100 V. High-frequency junction transistors for linear operations (e.g. buffering) on the signals generated in *p–n–p–n* circuits operate at lower voltage levels. Hence means are sometimes required for reducing the amplitude of *p–n–p–n* generator outputs below the striking voltage.

Consider the circuit of Fig. 8, which is identical with that of

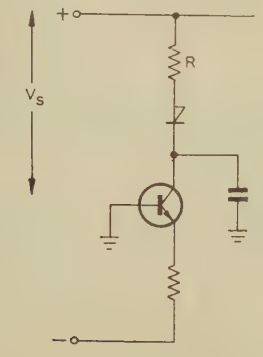


Fig. 8.—Transistor-controlled-*p–n–p–n* linear ramp generator producing reduced-amplitude output.

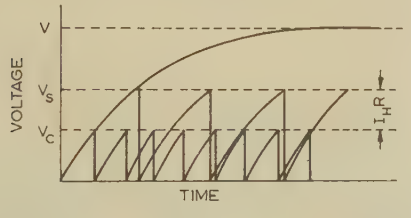


Fig. 9.—Dependence of output of circuit of Fig. 8 on series resistance *R*.

Fig. 5, except for the addition of a series resistance *R*. Fig. 9 shows the waveform of the functions generated. With zero series resistance the amplitude is nearly equal to the striking voltage, since the current through the *p–n–p–n* diode remains high until extinction takes place owing to reduction of the capacitor voltage below the turn-off voltage of the *p–n–p–n* diode. With finite series resistance *R*, the *p–n–p–n* diode 'extinguishes' when the current through the device becomes smaller than the holding current. This occurs at a capacitor voltage V_c such that $(V_s - V_c)/R \leqslant I_H$, where V_s is the striking voltage. The minimum voltage reached by the capacitor may be taken as close to zero, whence the ramp amplitude is nearly equal to V_c or $V_s - I_H R$. Hence variation of *R* permits control of amplitude over a wide range. Typically, the amplitude of a generator employing a 100 V *p–n–p–n* diode may be reduced to 20 V, which is well within the collector voltage rating of readily-available low-leakage silicon transistors. Furthermore, in the circuit of Fig. 8, reduction of ramp amplitude by means of series resistance reduces the maximum collector voltage applied to the control transistor, whence the collector leakage (see Appendix 11.2) is reduced. Very slow, substantially linear, ramps (typically one cycle in 500 sec) can therefore be generated, representing charging currents of about 1 μA into a few tens of microfarads. A practical circuit for producing very slow ramps and sine waves is shown in Fig. 15. The variation of amplitude from the circuit of Fig. 8, as *R* was varied from 47 Ω to 8 kΩ, is shown in Fig. 10.

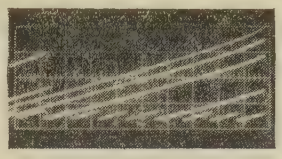


Fig. 10.—Variation of ramps generated by circuit of Fig. 8 as the series resistance was varied from 47 Ω (largest amplitude) through 1, 2 and 5 kΩ to 8 kΩ (smallest amplitude).
Horizontal scale: 100 μs/division.
Vertical scale: 20 V/division.

(4) VARIABLE-WIDTH PULSE CIRCUITS

The requirement sometimes arises for the generation of pulses having variable width. This is readily achieved, merely by resistance variation, using the *p–n–p–n* circuit of Fig. 11. Consider a fast triggering pulse applied to increase momentarily the

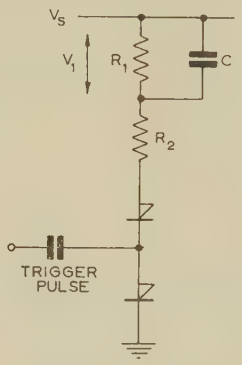


Fig. 11.—Variable-length pulse generator.

voltage on a supply line whose steady value is just smaller than the striking voltage V_s. The *p–n–p–n* diode commences to conduct a current whose initial magnitude is given by V_s/R_2, since, initially, the capacitor C is uncharged. The potential V_1 across R_1 and C rises exponentially according to

$$V_1 = V_s \frac{R_1}{R_1 + R_2}(1 - \varepsilon^{-t/\tau}) \quad . \quad . \quad . \quad . \quad (1)$$

(where τ is the time-constant produced by C, R_1 and R_2 in parallel) until $(V_s - V_1)/R_2 \leqslant I_H$, when the diode again leaves the conducting state. This condition is obtained from eqn. (1), after some manipulation, as

$$T = \tau \log_\varepsilon \left[\frac{R_1 V_s}{R_2(R_1 + R_2)I_H - V_s} \right] \quad . \quad . \quad (2a)$$

which approaches

$$T = CR_2 \log_\varepsilon \frac{V_s}{I_H R_2} \text{ when } R_1 \gg R_2 \quad . \quad . \quad . \quad (2b)$$

where *T* is the time during which the diode remains conducting, following a trigger pulse. Clearly, to produce a pulse of finite width, the current determined by V_s/R_2 must exceed I_H. Differentiation of eqn. (2*b*) with respect to R_2 shows that the maximum pulse length obtainable is equal to $0{\cdot}37 V_s C/I_H$, again

assuming that $R_1 \gg R_2$. Experimentally, pulse-width variations over the range 1–100 μs have been obtained merely by variation of R_2; variation of C, of course, permits control of pulse length over much wider ranges. The same circuit may be operated repetitively merely by raising the potential of the supply line slightly above the striking voltage V_s.

Applications of the variable-pulse-width circuit here described include generation of delay signals for oscilloscope circuits and time-interval generation generally. The accuracy of the circuit depends to a large extent on the constancy of the *p–n–p–n* diode 'hold' current. Shockley[4] has discussed the temperature dependence of this parameter; for *p–n–p–n* diodes having large 'hold' currents, say 30–50 mA, the variation may be as small as −0·25% per deg C. It is stated that the variation may be an order of magnitude larger for room-temperature 'hold' currents below about 5 mA.

(5) HARMONIC GENERATION

The short turn-on time of *p–n–p–n* diodes makes it possible to generate extremely sharp pulses. Their repetition rate can be readily controlled from external trigger signals, which may, for example, be pulses or sine waves derived from a quartz-crystal oscillator. Thus *p–n–p–n* circuits may be used to produce high-order harmonics up to several gigacycles per second. In this Section, the limits of harmonics which may be extracted under ideal conditions, i.e. purely resistive termination and no lead inductance, are considered. For the purpose of the analysis the limiting conditions for present *p–n–p–n* diodes are taken as

Maximum repetition rate*	1 Mc/s
Minimum switching time	1 ns (see Fig. 12)
Amplitude	100 V

* Repetition rates up to about 10 Mc/s are obtainable with avalanche transistors (see Section 7).

Taking the output of an idealized 1 Mc/s *p–n–p–n* generator to have the waveform indicated in Fig. 12, the frequency spectrum

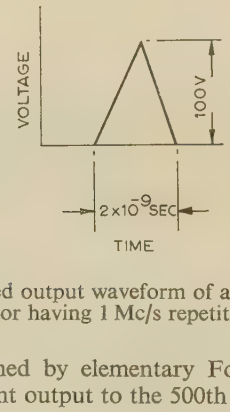


Fig. 12.—Assumed output waveform of avalanche harmonic generator having 1 Mc/s repetition rate.

of Fig. 13 is obtained by elementary Fourier analysis. This yields almost constant output to the 500th harmonic and energy in bands beyond the 1 000th harmonic. Taking the pulse height to be 100 V, the 500 Mc/s harmonic would have an amplitude of about 60 mV; the 1·5 Gc/s harmonic would be about 10 mV. These harmonics are of importance in the operation of sampling oscilloscopes,[5] as mentioned in Section 7 in connection with avalanche transistors. Moll, Uhlir and Senitsky[6] have reported radiation in the 9 Gc/s region from avalanching silicon diodes pulsed at a repetition rate of 500 kc/s.

(6) GENERATION OF SINE WAVES

Means for varying the repetition rate of linear ramp functions generated by *p–n–p–n* diode circuits were discussed in Section 2.

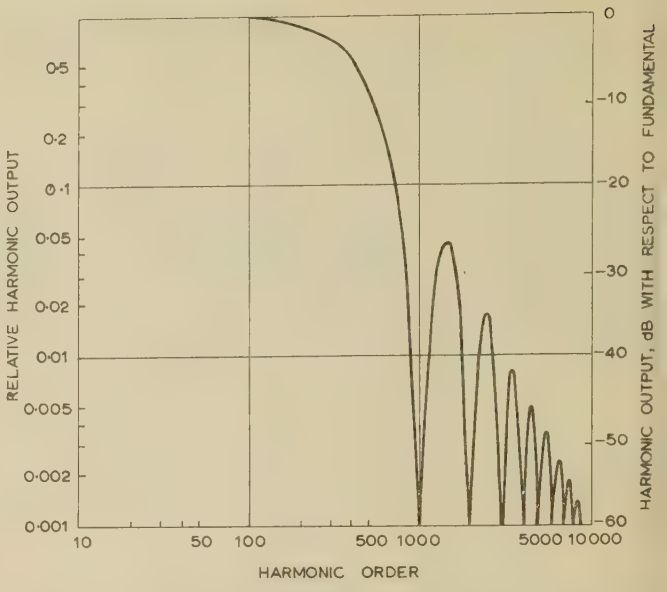


Fig. 13.—Computed spectrum for avalanche harmonic generator having 1 Mc/s repetition rate and 1 ns rise time (see Fig. 12).

These functions may readily be translated into sine waves thereby yielding very wide deviation frequency modulation.

Consider the function of Fig. 14(*a*), having zero d.c. component, applied to a bridge rectifier circuit. The output is

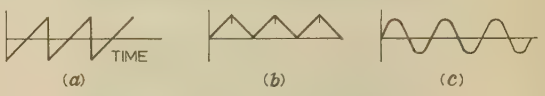


Fig. 14.—Steps in generation of sine wave.
(*a*) Ramp output from *p–n–p–n* generator.
(*b*) Rectified ramp output.
(*c*) Sine wave obtained after limiting and filtering.

triangular waveform [Fig. 14(*b*)], with a short transient corresponding to the finite switching time. Limiter circuits, followed by a low-pass filter to remove the switching transient, produce sinusoidal output at a frequency equal to the repetition frequency of the ramp function. As indicated in Fig. 5, this frequency can be made proportional to an input voltage. In particular, the input may itself be derived from a slow *p–n–p–n* generator, either linear or exponential, to achieve linear or logarithmic frequency sweeps over one or more decade ranges of frequency. Fig. 1 shows a practical circuit employed to generate sine waves having less than 5% non-linear distortion at frequencies down to one cycle in 5 min.

(7) AVALANCHE TRANSISTORS

The subject of avalanche transistors has been considered by Chaplin and Owens[7] and by Gage.[8] Fig. 16 shows an avalanche transistor relaxation oscillator. Operation of this circuit to repetition rate of about 10 Mc/s is possible with selected diffused base transistors, e.g. type 2N1143; alloy-junction transistors type OC44 have operated reproducibly to 4 Mc/s with 40 amplitude. The waveform generated in a time-base circuit of this type is shown in Fig. 17. Synchronization with waveforms having repetition rates to several hundred megacycles per second is readily achieved.

The switching time of avalanche transistors[7] may sometimes be in the sub-millimicrosecond range. The harmonic generation

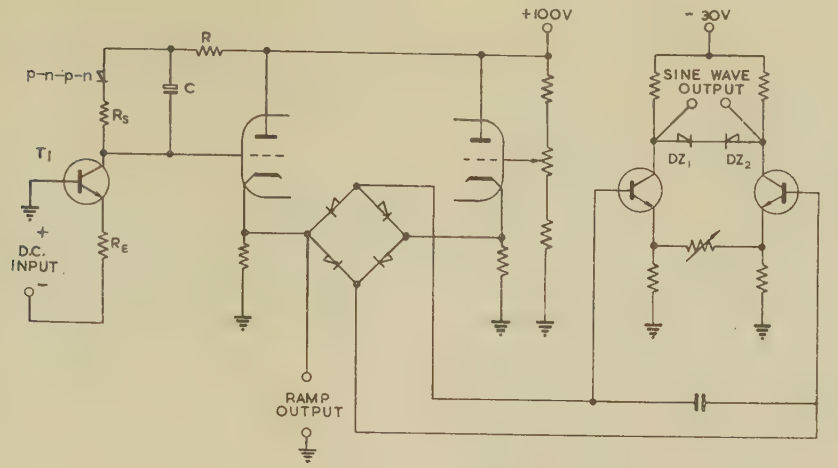


Fig. 15.—Low-frequency ramp and sine-wave generator operating from 1 cycle in 5 min to 1 cycle in 10 sec.

p–n–p–n transistor: DS1.
DZ: 10 V Zener diodes.
R_S: 47 Ω.
R: 3 000 Ω.
C: 50 μF tantalum capacitor.
T_1: *n–p–n* silicon transistor (type 2S003).

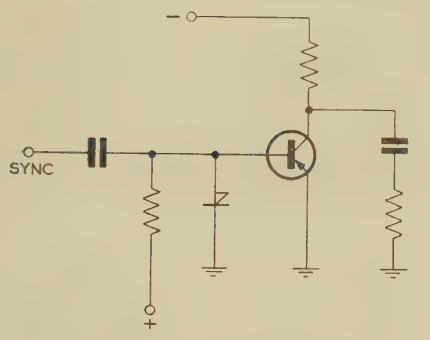


Fig. 16.—Avalanche-transistor ramp generator.

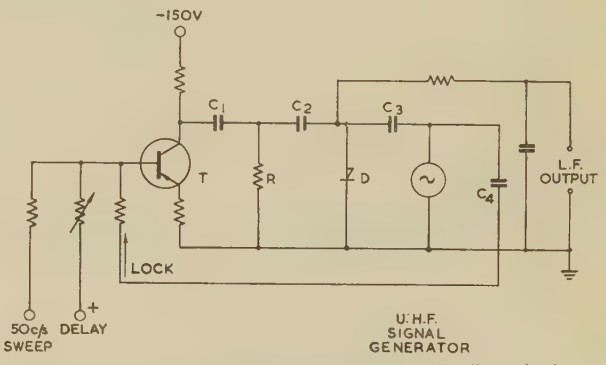


Fig. 18.—Single-transistor sampling head with direct lock.

T: avalanche transistor ASZ23.
D: diode CV448.
R: 220 Ω.
C_1, C_4: 22 pF.
C_2, C_3: 47 pF.

effects mentioned in Section 5 can, of course, also be achieved with avalanche transistors. Fig. 18 shows the circuit of a single-transistor sampling unit which has been tested up to 1·9 Gc/s. Oscillograms on a 50 c/s time-base (real time) obtained with sine-wave inputs ranging from 600 to 1 900 Mc/s are reproduced in Fig. 19. The output is obtained as a low-frequency beat

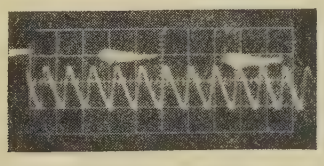


Fig. 17.—Output of 10 Mc/s avalanche-transistor ramp generator.

Transistor: type 2N1143.
Horizontal scale: 0·1 μs/division.
Vertical scale: 5 V/division.

between the unknown signal and a locked harmonic of the avalanche-transistor output. Phase modulation of the avalanche generator is obtained by applying to the avalanche-transistor base the output of a slow ramp generator or a signal derived directly from the a.c. supply mains. In view of the work reported by Goodall and Dietrich,[9] who obtained an equivalent sine-wave bandwidth of 5·5 Gc/s using vacuum-tube impulse generators, and of the work of Moll *et al.*[6] (referred to in Section 5), the 1·9 Gc/s response shown in Fig. 19 does not appear to be the maximum possible in avalanche-type sampling oscilloscopes.

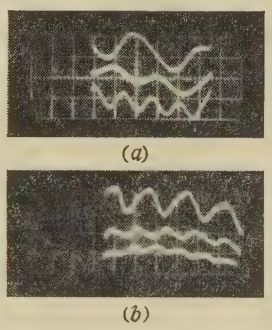


Fig. 19.—Output of sampling head with sinusoidal input.

(*a*) 0·6, 0·9 and 1·2 Gc/s; 250 mV input approximately.
(*b*) 1·3, 1·6 and 1·9 Gc/s; 500 mV input approximately.
Sensitivity of low-frequency display oscilloscopes: 5 mV/division.

(8) CONCLUSIONS

Applications of the *p–n–p–n* diodes to the generation of short pulses are discussed, and *p–n–p–n* ramp generators and digitizers capable of 0·1% linearity are described. Tests on triode transistors operated in the avalanche mode are mentioned; these have yielded repetition rates of 10 Mc/s and hold out the possibility of harmonic generation to several gigacycles per second with applications in sampling oscilloscopes.

(9) ACKNOWLEDGMENTS

Thanks are due to Marconi Instruments, Ltd., for permission to publish the paper, and to Messrs. O. H. Gigli, A. A. Smith and G. Barradell for carrying out some of the measurements described.

(10) REFERENCES

(1) SHOCKLEY, W.: 'Electrons and Holes in Semiconductors' (Van Nostrand, 1950), p. 112.
(2) MOLL, J. L., TANNENBAUM, M., GOLDEY, J. M., and HOLONYAK, N.: '*p–n–p–n* Transistor Switches', *Proceedings of the Institute of Radio Engineers*, 1956, **44**, p. 1174.
(3) ALDRICH, R. W., and HOLONYAK, N.: 'Multiterminal *p–n–p–n* Switches', *ibid.*, 1958, **46**, p. 1236.
(4) SHOCKLEY TRANSISTOR CORPORATION: Publication AD–1, December, 1958.
(5) CHAPLIN, G. B. B., OWENS, A. R., and COLE, A. J.: 'A Sensitive Transistor Oscillograph with D.C. to 300 Mc/s Response', *Proceedings I.E.E.*, Paper No. 2943, May, 1959 (**106** B, Suppl. 16, p. 815).
(6) MOLL, J. L., UHLIR, A., and SENITSKY, B.: 'Microwave Transients from Avalanching Silicon Diodes', *Proceedings of the Institute of Radio Engineers*, 1958, **46**, p. 1306.
(7) CHAPLIN, G. B. B., and OWENS, A. R.: 'A Method of Designing Avalanche Transistor Trigger Circuits', *Proceedings I.E.E.*, Paper No. 2944, May, 1959 (**106** B, Suppl. 16, p. 806).
(8) GAGE, D. S.: 'Pulse Circuits Using Avalanche Transistors', *Proceedings of the National Electronics Conference* (*Chicago*), 1958, **14**, p. 32.
(9) GOODALL, W. M., and DIETRICH, A. F.: 'Fractional Millimicrosecond Electrical Stroboscope', *Proceedings of the Institute of Radio Engineers*, 1960, **48**, p. 1591.

(11) APPENDICES

(11.1) **Effect of Finite Charging Time on Linearity of Avalanche Digitizer**

It is stated in Section 2 that the frequency of an avalanche digitizer should be proportional to the input voltage, provided that the voltage variation across the capacitor during the discharge period is linear. The non-linearity introduced into the digitizer characteristic due to finite charging time is discussed in this Appendix.

Consider the equivalent circuit of Fig. 20(*a*), showing a

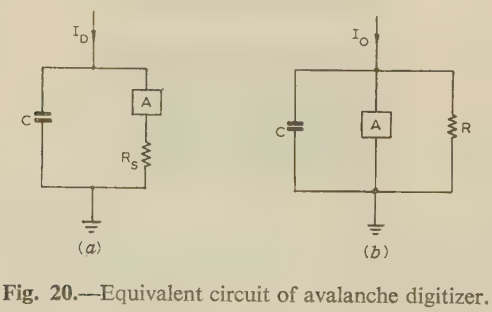


Fig. 20.—Equivalent circuit of avalanche digitizer.
(*a*) With finite charging time.
(*b*) With capacitor or transistor leakage.

capacitor C discharged by a constant current I_d, connected in parallel with an avalanche device A breaking down at a voltage V_s from a high-resistance to a low-resistance state. The series resistance R_s in Fig. 20(*a*) includes the effective 'on' resistance of the avalanche device and any additional current-limiting resistance. The time taken for a voltage change V_s when a constant current I_d is taken from the capacitor is, of course, given by the relation $t_d = CV_s/I_d$. The charging time t_c depends to some extent, on the 'hold' current of the avalanche device according to

$$I_H = \frac{V_s}{R_s} \exp\left(-\frac{t_c}{R_s C}\right)$$

since charging can only continue from an initial current V_s/R_s until the current has fallen to the hold current I_H. Evidently the ratio t_c/t_d for the maximum frequency of operation is a measure of the non-linearity of the digitizer characteristic.*

From the expressions given above for t_c and t_d, their ratio can readily be seen to be

$$\frac{t_c}{t_d} = \frac{I_s R_s}{V_s} \log_e \frac{V_s}{I_H R_s} \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Taking typical values for the parameters of eqn. (3), as follows:

$V_s = 100\,\text{V}$.
$I_H = 1$–$10\,\text{mA}$ at maximum frequency of operation.
$V_s/R_s = 5\,\text{A}$ (determined by maximum current rating of the avalanche device).
$R_s = 20\,\Omega$.
$I_H = 1$–$50\,\text{mA}$,

the values of t_c/t_d shown in Table 1 are obtained. The Table shows t_c/t_d for 'hold' currents of 1, 5 and 50 mA, and for dis

Table 1

COMPUTED DEPENDENCE OF t_c/t_d RATIO* ON DISCHARGE AND HOLD CURRENTS

I_c	t_c/t_d			f_{max}
	$I_H = 1\,\text{mA}$	$I_H = 5\,\text{mA}$	$I_H = 50\,\text{mA}$	
mA				kc/s
1	0·001 6	0·001 4	0·001	1
2	0·003 3	0·002 8	0·002	2
5	0·008	0·007	0·005	5
10	0·016	0·014	0·01	10

* Percentage non-linearity of digitizer characteristic is given by $100t_c/t_d$.

charge currents (which determine maximum frequency of operation with a given capacitance) of 1, 2, 5 and 10 mA. Maximum operating frequencies, for the arbitrary case of 0·01 μF timing capacitance, are also given.

It is seen that non-linearity does not exceed 0·2–0·3% for any 'hold' current where a digitizer is operated at a maximum discharge current of 2 mA. Effects of shunt resistance (capacitor leakage and finite transistor collector resistance) are considered in detail in Appendix 11.2. These are unlikely to produce a leakage current greater than 2 μA for silicon control transistors operated with high emitter resistance, as indicated in Fig. 20(*b*). Taking again a maximum discharge current of 2 mA, the leakage represents an error of 0·1% at maximum frequency. Thus a ratio of maximum to minimum frequency of the order of 1 000 : 1 appears possible. The experimental characteristic in Fig. 6 represents a ratio of 500 : 1.

(11.2) **Relation between Ramp Linearity and Dynamic Range of Avalanche Digitizer**

In avalanche digitizers a limitation to operation over extremely wide frequency ranges is imposed by leakage due to the fir

* In this analysis, the intrinsic switching time for the avalanche device is ignored. This is usually negligible compared with the time period ($t_c + t_d$) of avalanche digitizers operating up to some tens of kilocycles per second only.

collector resistance of the control transistor and by capacitor leakage. These quantities can be represented by an equivalent shunt resistance across the timing capacitor.

In the equivalent circuit of Fig. 20(*b*), a leakage resistance R is connected across a timing capacitance C shunted by an avalanche device breaking down at V_s volts. It is evident that breakdown occurs, and ramp waveforms are generated, only if $I_0R > V_s$, and operation ceases when $I_0R = V_s$. Thus, if a maximum input current nI_0 is supplied (corresponding to the current I_d considered in Section 11.1 and usually determined by maximum power dissipation in the control transistor), there will be a maximum linearity error of $1/n \times 100\%$.

For the same circuit, when operating at an input current nI_0, the non-linearity of the ramp function generated is determined by the slopes of the exponential

$$V = nI_0R[1 - \exp(-t/RC)]$$

for $V = 0$ and $V = V_s$. The two slopes are given by

$$\frac{dV}{dt}_{V=0} = \frac{nI_0}{C} \text{ and } \frac{dV}{dt}_{V=V_s} = \frac{nI_0}{C}\left(1 - \frac{V_s}{nI_0R}\right)$$

whence the ratio of final and initial slopes equals $1 - V_s/nI_0R$ and the amount of non-linearity follows as $(V_s/nI_0R) \times 100\%$. Taking I_0, as before, to equal V_s/R (the minimum current needed for repetitive operation) the percentage non-linearity at maximum input may be written as $1/n \times 100$. This is the same expression as that for the maximum linearity error of the voltage/frequency curve.

The above method for determining ramp linearity, by noting the range of the digitizer, has been found useful, since it is not readily possible to measure directly non-linearity of ramp functions to, say, 0·1%. Using the voltage/frequency curve, however, measurements are made merely in terms of steady-state voltage and frequency, whence high accuracy is possible.

621.382.3

The Institution of Electrical Engineers
Paper No. 3671 M
Sept. 1961

SOME ADVANTAGES OF SILICON TRANSISTORS IN CIRCUIT DESIGN

By M. K. McPHUN, B.Sc.(Eng.), Graduate.

(*The paper was first received* 10*th March, and in revised form* 17*th May*, 1961.)

SUMMARY

After transistors had been used for industrial instrumentation within the Development and Engineering Group of the United Kingdom Atomic Energy Authority for some years, silicon transistors were adopted for general-purpose use, as they were expected to be more reliable than germanium transistors. It has since become apparent that they possess more advantages over germanium transistors than is generally recognized, or was thought at the time of their adoption. For example, direct coupling of transistor circuits is facilitated and the number of components required may be much reduced; the use of electrolytic capacitors may be avoided.

The paper compares the performance of silicon and germanium transistors from the viewpoint of the circuit designer; physical origins of their characteristics are not discussed. A range of circuits which take advantage of the characteristics peculiar to silicon transistors is described. The circuits are for direct-coupled amplifiers, current amplifiers for small signals, switching applications and multivibrators with long periods.

(1) INTRODUCTION

Since the advent of silicon transistors, descriptions have been given of their use for special purposes, e.g. choppers[1] and low-drift input stages for d.c. amplifiers.[2]

Reliability is of prime importance in industrial control equipment, particularly in the atomic energy industry. As soon as an assured supply of silicon transistors became available, it was decided to use them instead of germanium transistors for general purposes in the Central Instrument Laboratory of the U.K.A.E.A. (Development and Engineering Group) for the following reasons:

(*a*) They were expected to be more reliable than germanium transistors.
(*b*) They could be operated at higher temperatures than germanium transistors, a particularly useful feature in chemical plants.
(*c*) The manufacturers predicted that their costs would fall below those of germanium transistors within a few years, making germanium transistors obsolescent for general purposes.

The paper deals with experience gained in using silicon transistors for a wide variety of applications. The design of the individual circuits for these applications is not described in full where the techniques used are conventional; only those points of special relevance to the use of silicon transistors are described.

(2) COMPARATIVE PERFORMANCE OR GERMANIUM AND AVAILABLE SILICON TRANSISTORS

This discussion is concerned with transistor performance as it affects the circuit designer; the physical origins of the characteristics are not considered.

(2.1) Input Characteristics

The current/voltage curves for the emitter-base diode of typical germanium and silicon alloy-junction transistors (Fig. 1) show a sharper turnover for the silicon than for the germanium transistor. As with the forward-biased silicon diode, the forward voltage across the emitter-base diode may be considered as constant at 0·6 V. For applied voltages of less than 0·4 V the junction may be considered to be reverse-biased. This forward bias necessary for conduction may not be ignored in circuit design as often as it is with germanium transistors; it may, however, be put to good use, as is shown throughout the paper.

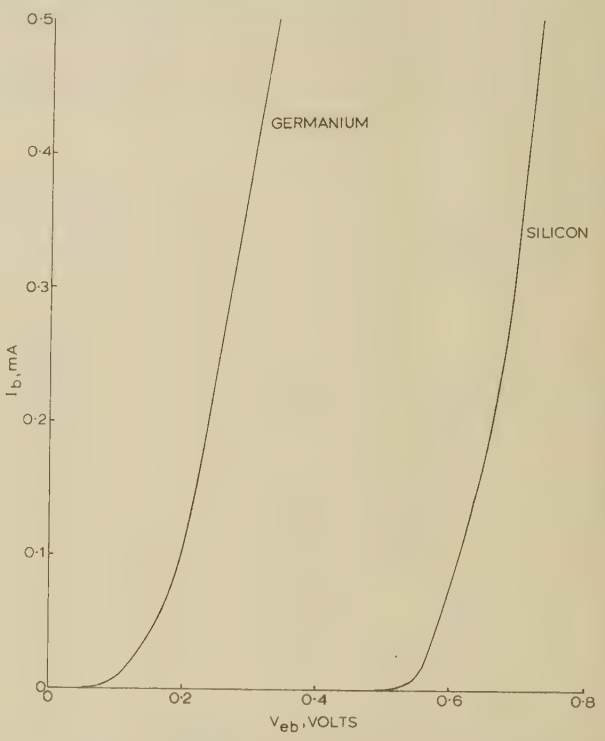


Fig. 1.—Typical characteristics of emitter-base diodes for alloy-junction germanium and silicon transistors.

(2.2) Leakage Current and Temperature

The collector current flowing with open-circuit emitter is much smaller for the silicon than for the germanium transistor. This is the common-base leakage current, I_{co}; the published applications[1, 2] of silicon transistors make use of the fact that I_{co} is very small so that its dependence on temperature has much less effect on circuit performance than with the germanium transistor. The effect of temperature on the leakage current is the same for both germanium and silicon transistors, i.e. an increase in temperature of about 9° C causes the leakage current to double; Table 1 shows typical figures for silicon and germanium alloy-junction transistors.

In many general-purpose circuits the leakage current of the silicon transistor may be completely neglected, thus obviating the need for stabilizing the d.c. working point against changes in leakage current with temperature; this leads to simpler circuits and greater reliability.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Mr. McPhun was formerly with the United Kingdom Atomic Energy Authority (Development and Engineering Group) and is now at the Central Electricity Research Laboratories.

Table 1

COMPARISON OF LEAKAGE CURRENTS AT 25° C FOR GERMANIUM AND SILICON TRANSISTORS

Description	Type	I_{co}	I'_{co}
		μA	μA
A.F. germanium	OC71	5	150
R.F. germanium	OC44	0·5	25
Silicon	OC201	10^{-3}	60×10^{-3}

The other main temperature effect is the variation of V_{cb} with temperature at constant collector current. This effect is the same for both silicon and germanium transistors, being about 2·5 mV per deg C.

The permissible range of operating temperatures is much greater for silicon than for germanium transistors. Typical maximum operating temperatures are germanium, 75° C and silicon, 150° C.

(2.3) Collector Resistance

Measured values of common-base collector resistance, r_c, at normal operating voltages of the order of 6 V are lower for typical alloy-junction silicon transistors than for similar germanium types, but they are markedly higher at low values of collector-base voltage, V_{cb}.

The characteristic of the collector-base diode of the simple equivalent circuit shown in Fig. 2 is similar to that of the

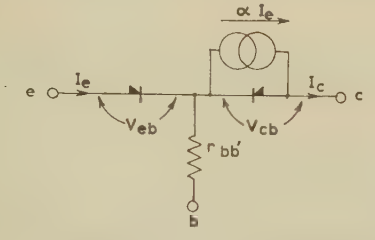


Fig. 2.—Simple equivalent circuit of transistor.

emitter-base diode, Fig. 1. For normal operation as an amplifier, the emitter-base diode is forward-biased and the collector-base diode is reverse-biased. However, the silicon diode is effectively reverse-biased for voltages through zero to forward values of a few tenths of a volt, and it would thus be expected that transistor action would be maintained with high collector resistances for collector-base voltages in this range. This is indeed so, as shown in Fig. 3 which gives measured curves of r_c plotted against V_{cb} for both germanium and silicon transistors. Over the range 0·2 V reverse to 0·1 V forward for V_{cb}, the collector resistance of the silicon transistor decreases by one-third while that of the germanium transistor decreases by three decades.

(2.4) Noise

Measurements[3] showed that the best type OC200 silicon transistors were slightly less noisy than the best available germanium transistors. The manufacturers claim the silicon OC201 to have a lower noise figure than the OC200, and measurements show a noise figure of 6 dB with a source resistance of 500 Ω to be easily obtainable at 1 kc/s with the OC201.

The OC201 was adopted for general-purpose use, and experience has shown that its noise figure is comparable with that of the best available germanium transistors.

A lower noise figure is obtained by operating a transistor at very low values of collector current.[3] This is facilitated with silicon transistors, as collector currents of 50 μA or less may be used in low-noise circuits without I'_{co} proving troublesome; under these conditions linearity may be a limiting factor.

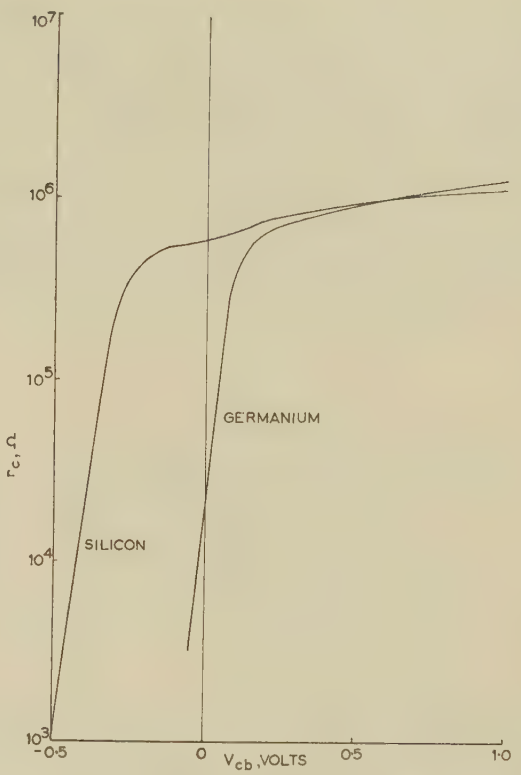


Fig. 3.—Collector resistance as a function of collector-base voltage for typical alloy-junction germanium and silicon transistors.
Collector current = 1 mA.

(2.5) Availability of Silicon Transistors and Quantitative Data

At present there are few types of silicon transistors available in this country;* all dissipate about 200 mW at room temperature and have common-base cut-off frequencies between 1 and 4 Mc/s. The germanium transistors of similar size and performance dissipate only 100 mW at room temperature because of their lower maximum junction temperature.

Silicon power and high-frequency transistors are now becoming available, but their price restricts their use to special applications. The cut-off frequency of the alloy-junction silicon transistors, however, has been found adequate for most instrument work, the exception being due to the difficulty in making stable wide-band amplifiers with large amounts of overall negative feedback.

So far a very large spread in the characteristics of silicon transistors must be tolerated; values of the common-emitter current gain α' from 20 to 130 may be expected in a batch of 20 transistors. New types with a narrow spread of α' have recently become available, but even with these, values of α' may still be found in the range 30–75.

Quantitative information about silicon transistors is scarce, and this puts circuit designers at a disadvantage. Most characteristics must be measured by the user, and because of the large spread this can be a very laborious process. For this reason, most of the circuits described here will tolerate a very wide spread of transistor characteristics.

(3) DIRECT-COUPLED AMPLIFIERS

The frequent requirement for 50 c/s servo amplifiers showed the need for a.c. amplifiers using direct coupling between stages;

* A rapid increase in the variety of silicon transistors available has taken place since the paper was written.

electrolytic capacitors were usually required for interstage coupling because of the low impedance of transistor circuits. Greater reliability should be expected from amplifier circuits using no electrolytic capacitors, and any reduction in the number of components used should also increase reliability.

Amplifiers consisting of common-emitter stages with each collector directly coupled to the base of the succeeding transistor have been suggested,[4, 5] and practical examples of this type of circuit using germanium transistors have been described.[1, 4, 5]

Considerable advantages are gained by using silicon transistors in direct-coupled amplifiers; the basic form of the circuit is shown in Fig. 4. Transistors T_1 and T_2 are operating with

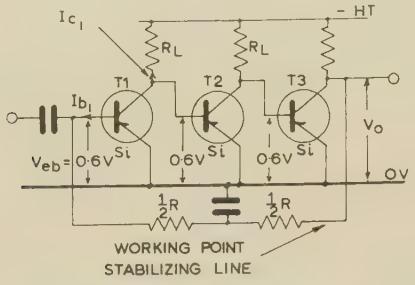


Fig. 4.—Basic direct-coupled a.c. amplifier.

$V_{cb} = 0$, a condition shown clearly in Fig. 3. Whereas the stage gain with germanium transistors is very low because of the low value of r_c at $V_{cb} = 0$, that with silicon transistors is normal for the case where $r_c(1 - \alpha) > R_L$.

Let us consider germanium transistors used in this circuit.[5] The collector resistance is reduced to one-hundredth of the value obtained with a few volts applied to the collector, then $r_c(1 - \alpha) \ll R_L$, and this severely limits the gain per stage; moreover, under these conditions r_c is very sensitive to temperature, and so also is the stage gain. The voltage swing at a collector must be limited to the millivolt region because of the change in r_c with V_{cb}. A voltage-divider coupling must be used to the last stage[4] in order to allow sufficient voltage swing at the base of T_3; this, in turn, means that a positive supply rail must be used.

With silicon transistors, a high value of r_c is maintained, and so $R_L \ll r_c(1 - \alpha)$; the stage gain is not then limited by r_c and the small variations in r_c with V_{cb} have negligible effect. The maximum voltage swing at a collector may be of the order of 200 mV; this is sufficient to provide a signal at the T_3 collector with a peak voltage of a few volts, no positive rail being required for an interstage coupling. Further, as the stage gain is governed by R_L and not r_c, any temperature-dependence of r_c does not affect the gain.

Table 2

EFFECT OF GAIN OF T_1 ON V_o (FIG. 4)

α_1'	I_{b1}	V_o
	mA	V
30	1/30	10
40	1/40	7·5
60	1/60	5

The simple operating-point stabilization shown in Fig. 4 is adequate for most purposes. The required base direct current, I_{b1}, of the first stage is determined; R is then chosen so that $V_o - V_{eb} = I_{b1}R$. Variations of I_{co}' with temperature may be neglected as this current is so small; variations of V_{eb} with temperature are much less than V_o. Difficulty is encountered, however, if it is desired to have a close control on V_o, and at the same time to accommodate transistors demanding widely differing values of I_{b1}.

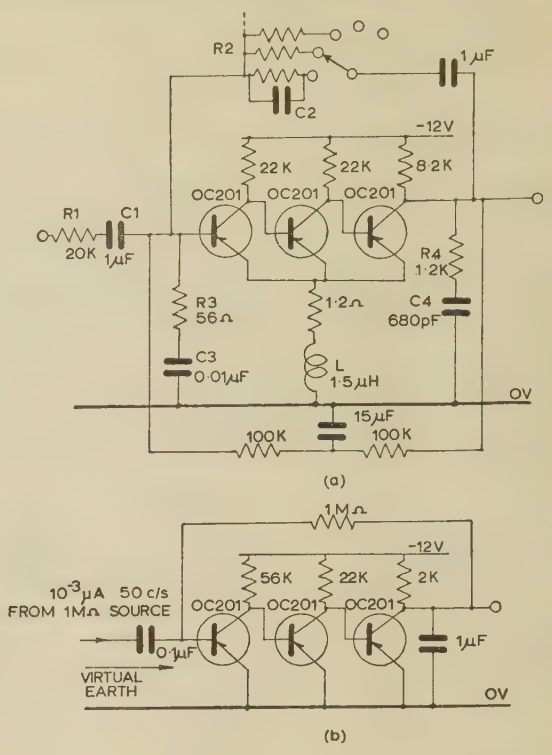


Fig. 5.—Practical examples of direct-coupled amplifiers.
(*a*) Wide-band amplifier for use with 1kc/s square waves.
(*b*) 50c/s chopper amplifier.

The solution adopted here is to have a close control on the value of α_1' for the first transistor T_1. For example, with $I_{c1} = 1$ mA, $R = 300$ kΩ, the values shown in Table 2 will be obtained. The variations of α_1' tolerated will be governed by the permissible variation of V_o.

Practical examples of these circuits are shown in Fig. 5. Fig. 5(*a*) shows a wide-band amplifier for use with 1 kc/s square waves. Fixed current feedback is applied, and gains of 1, 2, 5, 10, 20 and 25 are obtained by switching the shunt feedback resistors. Stabilizing networks R_3C_3, R_4C_4, C_2 and L are necessary because of the large degree of feedback and wide bandwidth of 100 kc/s.

For the low-noise chopper amplifier shown in Fig. 5(*b*) it was convenient to use the same resistor for feedback at the 50 c/s signal frequency as for stabilizing the d.c. working point. An electrolytic capacitor would have been required for decoupling a separate working-point stabilization loop at 50 c/s.

(4) CURRENT AMPLIFIERS FOR SMALL SIGNALS

It is sometimes inconvenient to use an electrometer valve in conjunction with equipment otherwise using transistors, usually because the power supplies required are not available. Current amplifiers using silicon transistors have been successfully used in such applications and two examples are described.

Fig. 6 shows the circuit of a head amplifier for a photomultiplier detecting very low levels of light chopped at 1 kc/s. Two cascaded emitter followers are used, giving a current gain of about 300. The collector current of T_1 is set at 120 μA. This is a compromise between the low value required for low noise and the need for linearity over a range of input currents of three decades, from 10^{-8} to 10^{-11} A peak. The noise generated in the first transistor is negligible compared with the shot noise of

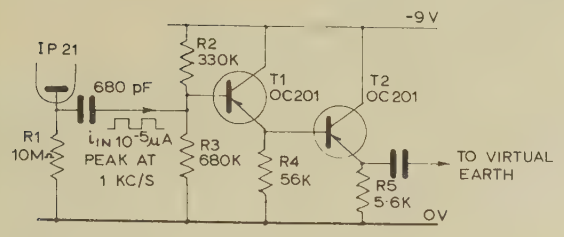


Fig. 6.—Circuit of photomultiplier head amplifier.
The type IP21 photomultiplier is selected for low noise.

the photomultiplier. It is possible to observe individual electrons leaving the photomultiplier cathode; this suggests that the circuit could be used with counting equipment for photon counting at very low light levels.

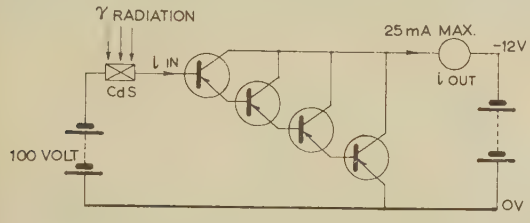


Fig. 7.—Circuit of amplifier for crystal-type radiation detector.
Transistors are type OC201.

Fig. 7 shows the circuit of an amplifier for direct currents suitable for use with crystal radiation detectors. A typical cadmium-sulphide detector in this circuit will pass current, i_{in}, up to 2 μA. The minimum current gain is $1{\cdot}3 \times 10^7$. The total leakage current at the output is 1·3 μA at 25° C; thus, at room temperature, operation over a range of four decades is possible for values of i_{out} from 2·5 μA to 25 mA. The linearity over this range is only 10% but this is unimportant in this application as the crystal radiation detectors are non-linear.

(5) SWITCHING APPLICATIONS

Since a forward bias of 400 mV is required at the base of a silicon transistor to start conduction, it is easy to ensure that it does not conduct; it is never necessary to make the base positive with respect to the emitter to reduce the collector current to below 1 μA at room temperature, and thus no positive supply rail is required.

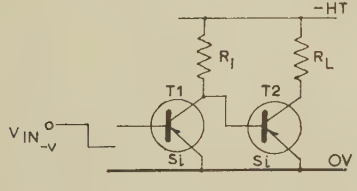


Fig. 8.—Direct-coupled switching circuit.

In the circuit of Fig. 8, T_2 may be adequately cut off by bottoming the collector of T_1; there is a clear margin between the bottoming voltage of T_1, typically 100 mV, and the 450 mV required at the base of T_2 for conduction.

(5.1) Relay-Drive Circuits

A relay may be substituted for R_L in Fig. 8, if protection is provided for T_2 against voltage overshoot when T_2 is being cut off.[6] Apart from ensuring that transistor ratings are not exceeded, the only design requirement is that the current through R_1 with T_1 cut off is sufficient to bottom T_2.

If sufficient voltage swing is available to drive the relay and current amplification only is required, the circuit of Fig. 9 may

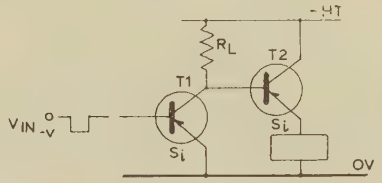


Fig. 9.—Relay-drive circuit.

be used. With T_1 cut off, T_2 is saturated. When T_1 is bottomed, T_2 will continue to conduct at first as the stored energy of the relay when released causes conduction of the emitter-base junction. Thus, no surge protection is required for T_2 but the relay is necessarily slow to release.

When it is desirable to operate T_1 with the minimum possible collector current, a more efficient circuit is obtained by using all the collector current of T_1 as base current for T_2. The circuit of Fig. 10 does this, using an *n–p–n* transistor for T_2.

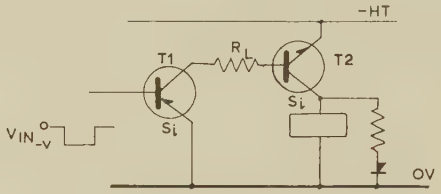


Fig. 10.—Efficient relay-drive circuit.

Both T_1 and T_2 are cut off simultaneously and bottomed simultaneously. The value of R_L fixes the T_1 collector current when T_1 is bottomed, this being made equal to the base current required to bottom T_2.

(6) MULTIVIBRATOR WITH LONG PERIOD

It is now generally recognized that a better astable multivibrator is obtained by placing the timing circuits in the emitters instead of in the collectors of transistors[7, 8]; when both junctions are reverse-biased the emitter leakage current is much less than the collector leakage current and I_{co} is the sum of these two.[9] However, there is an exception when long periodic times are required, since the emitter-timed circuit requires larger timing capacitances than does the base-timed circuit. To avoid the use of electrolytic capacitors for the timing circuits it is necessary to keep the timing capacitance as small as possible.

The small leakage current of silicon transistors allows the use of the base timing circuit, and also high impedances. It is shown in Section 10 that the maximum usable base timing resistance is obtained by returning it to an infinite aiming potential, V (Fig. 11). Little is gained by increasing V beyond $2V_{HT}$, and unless the additional supply rail is available, V will be made equal to V_{HT}. Fig. 11 also gives the component values required for a symmetrical astable multivibrator with a periodic time of 2 sec. The collector current when bottomed is only 100 μA, requiring a base current of 5 μA.

An asymmetrical circuit used to drive a tape reader via a relay at a frequency of 0·5 c/s is shown in Fig. 12. An asymmetrical circuit was needed to provide sufficient base current to the relay-drive transistor, T_3. The relay-drive circuit of Fig. 10

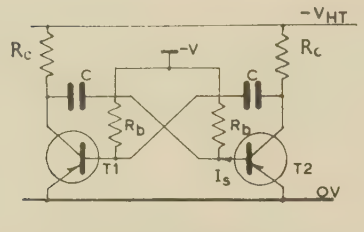


Fig. 11.—Basic multivibrator circuit.
The circuit is a symmetrical 0·5 c/s multivibrator with
R_c: 120 kΩ $V = V_{HT} = -12$ V
R_b: 2·2 MΩ T_1, T_2: OC202
C: 0·6 μF

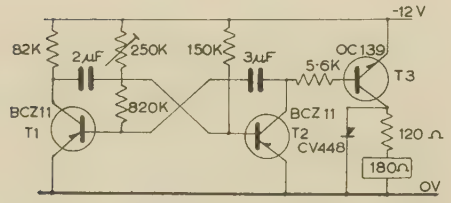


Fig. 12.—Asymmetrical multivibrator for driving a tape reader.

was used to give the greatest economy of current in T_2. A germanium transistor was used for T_3 as suitable silicon *n–p–n* transistors were not available at the time, although they are now becoming available in this country. The collector current in T_1 when bottomed is 140 μA; that in T_2 is 2 mA.

(7) CONCLUSIONS

Silicon transistors have been successfully applied to general-purpose use in industrial instrumentation to give increased reliability; none has yet failed in use in the experience of the laboratory.

The use of silicon transistors has led to simpler circuits with fewer components than if germanium transistors had been used, and the process of circuit design has been simplified. In particular, the use of electrolytic capacitors has been avoided, leading again to increased reliability.

The predicted decrease in cost of silicon transistors has occurred, to the extent of about one-half.

(8) ACKNOWLEDGMENTS

Mr. D. F. Davidson, of the Central Instrumentation Laboratory, Capenhurst, was responsible for introducing silicon transistors for general-purpose use in the laboratory and suggested their use in direct-coupled circuits. Helpful discussions were held with Mr. G. McGonigal of the same laboratory, and Mr. H. Jones developed the circuit shown in Fig. 7.

The paper is published by permission of Sir William Cook, Managing Director, and Dr. H. Kronberger, Deputy Managing Director, of the United Kingdom Atomic Energy Authority (Development and Engineering Group).

(9) REFERENCES

(1) Hutcheon, I. C., and Summers, D.: 'A Low-Drift Transistor Chopper-Type D.C. Amplifier with High Gain and Large Dynamic Range', *Proceedings I.E.E.*, Paper No. 3227 M, March, 1960 (**107** B, p. 451).
(See also Nambiar, K. P. P.: Discussion on the above paper, p. 461.)

(2) Kemhadjian, H.: 'Transistor Amplifiers for D.C. Signals', *Mullard Technical Communications*, 1958, **4**, p. 162.

(3) Battye, C. K., and George, R. E.: 'Transistors as Low Noise Amplifiers', *Proceedings I.E.E.*, Paper No. 2933 E May, 1959 (**106** B, Suppl. 18, p. 1190).

(4) Chaplin, G. B. B., Candy, C. J. N., and Cole, A. J. 'Transistor Stages for Wide-Band Amplifiers', *ibid.* Paper No. 2892 E, May, 1959 (**106** B, Suppl. 16, p. 762)

(5) Chaplin, G. B. B., and Owens, A. R.: 'A Transistor High-Gain Chopper-Type D.C. Amplifier', *ibid.*, Paper No. 2442 M, November, 1957 (**105** B, p. 258).

(6) Hill, C. F.: 'Transistors in Relay Switching Circuits' *Mullard Technical Communications*, 1956, **2**, p. 284.

(7) Bowes, R. C.: 'Timing Circuits and Waveform Generators' *R.R.E. Journal*, April, 1960, No. 44, p. 121.

(8) Bowes, R. C.: 'A New Linear Delay Circuit based on an Emitter-Coupled Multivibrator', *Proceedings I.E.E.*, Paper No. 2951 E, May, 1959 (**106** B, Suppl. 16, p. 793).

(9) Ebers, J. J., and Moll, J. L.: 'Large-Signal Behaviour of Junction Transistors', *Proceedings of the Institute of Radio Engineers*, 1954, **42**, p. 1761.

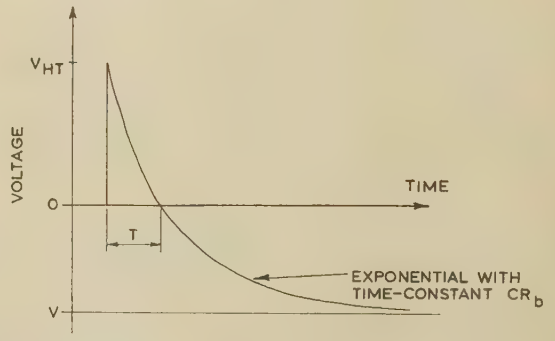


Fig. 13.—Timing action of multivibrator of Fig. 11.

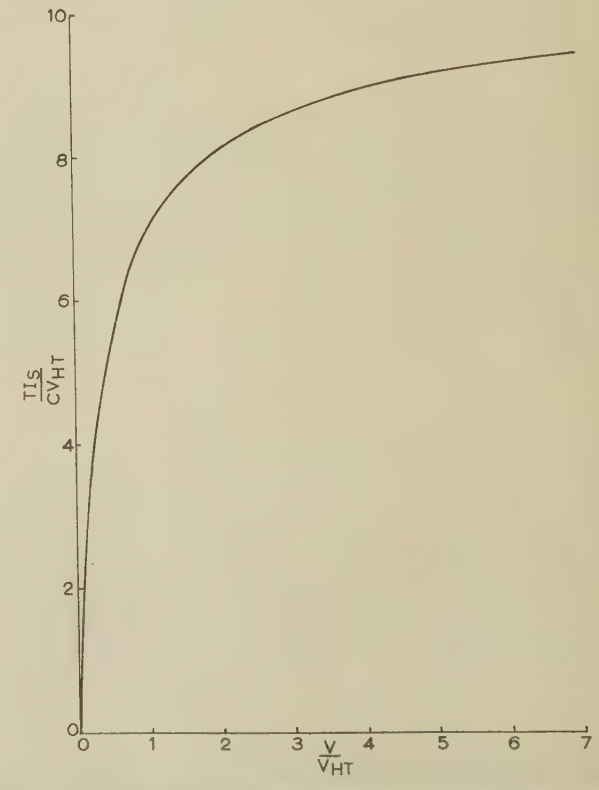


Fig. 14.—Calculated curve of TI_S/CV_{HT} as a function of V/V_{HT}.

(10) APPENDIX: MAXIMUM TIMING RESISTANCE FOR A MULTIVIBRATOR WITH LONG PERIOD

Referring to the circuit of Fig. 11, we require the value of V to give the maximum value of R_b. The current I_S required to bottom the transistor will be determined by R_b and V, which both enter also into the expression for the timing operation.

Let the time that a transistor is cut off be T. The timing action is shown in Fig. 13. The bottoming voltage is neglected and the transistor is assumed to conduct as soon as its base goes negative.

It is known that $T = CR_b \log_\varepsilon \left(1 + \frac{V_{HT}}{V}\right)$

We have also $R_b = \frac{V}{I_S}$

Eliminating R_b, we obtain

$$T = \frac{CV}{I_S} \log_\varepsilon \left(1 + \frac{V_{HT}}{V}\right)$$

In terms of V_{HT}/V as a variable,

$$\frac{TI_S}{CV_{HT}} = \frac{V}{V_{HT}} \log_\varepsilon \left(1 + \frac{V_{HT}}{V}\right)$$

In Fig. 14 is shown a dimensionless plot of TI_S/CV_{HT} against the ratio V/V_{HT}. Little is gained by increasing V beyond $2V_{HT}$.

MONOGRAPHS PUBLISHED INDIVIDUALLY

Summaries are given below of monographs which have been published individually, price 2s. each (post free). Applications, quoting the serial numbers as well as the authors' names, and accompanied by a remittance, should be addressed to the Secretary. For convenience, books of five vouchers, price 10s., can be supplied.

Intermodulation Noise in Linear F.M. Systems. Monograph No. 459 E.

R. I. MAGNUSSON, Tekn.Lic.

By applying a series of detailed approximations to the Rice and Bosse theory of f.m. intermodulation noise, the recent formulae of Medhurst and Roberts for intermodulation noise/signal ratios in f.m. systems with white-noise modulation and networks of polynomial frequency characteristics are extended in three respects: (*a*) to the calculation of the noise ratio at any modulation frequency, (*b*) to network real and imaginary characteristics of eighth and seventh degree, respectively, and (*c*) to the case of pre-emphasis. The theory is illustrated by application to the cases of a signal with C.C.I.R. pre-emphasis passing a stagger-tuned triple without and with group-delay equalization.

A Proof of the Generalized Topological Kirchhoff's Rules. Monograph No. 460 E.

A NATHAN, E.E., M.S., D.Sc.(Eng.).

Kirchhoff's topological rules for linear reciprocal networks have been generalized for networks containing non-reciprocal elements as well, in order to be applicable to active networks. A conventional proof of these rules for reciprocal networks is given, and the generalized rules are then shown to be an immediate consequence of this proof, thus requiring no special considerations for their derivation. The exposition is restricted to nodal analysis.

Theoretical and Experimental Analysis of the Cavity Maser. Monograph No. 461 E.

P. HLAWICZKA, B.Sc.(Eng.).

An exact analysis of the gain function of the cavity maser is made and is presented in graphical form suitable for correlation with experiment. Combined with measurements of the reflection coefficient of the passive cavity, and gain and bandwidth of the maser, the results yield accurate values for the unloaded, external and magnetic Q-factors of the cavity maser. The analysis also applies to any other resonant one-port device, e.g. a parametric amplifier.

A Loop Antenna Coupled to a Four-Wire Line and its Possible Use as an Element is a Circularly Polarized End-Fire Array. Monograph No. 462 E.

KUN-MU CHEN, Ph.D., and RONOLD W. P. KING, Ph.D.

A loop antenna coupled electromagnetically to a four-wire transmission line is studied theoretically and experimentally in terms of two mutually perpendicular two-wire lines that carry currents that are 90° out of phase. The size of the loop at which resonance occurs is determined. The zero-order and first-order solutions of the current distribution are evaluated, and it is found that at resonance the induced currents in the loop due to each two-wire line oscillate in a dipole mode of the loop. The currents induced by the four-wire line circulate as a travelling wave. The radiation field pattern is also evaluated. The square loop is studied first; it is later shown that the circular loop closely resembles the square loop. An array that consists of many loop antennae coupled electromagnetically to a four-wire transmission line may be expected to have a very high gain with circularly polarized radiation.

A Direct Procedure for the Synthesis of Resistive $(n + 1)$-Poles. Monograph No. 464 E.

Dr. P. P. CIVALLERI

The paper deals with the problem of the realization of a real nth-order square symmetrical matrix as the conductance matrix of a resistive n-port based on a network of $n + 1$ nodes.

The technique of synthesis starts with the construction of a tree of the independent voltages along well-known lines, by inspecting the sign pattern in the given matrix. It is shown that, at each stage of this realization process, one can also determine the branch conductances which have to be connected in order to fulfil the requirements of the given matrix, thus entirely realizing the synthesis procedure.

The advantage of this method is that it does not require long algebraic computations in the final steps of the synthesis.

The branch-conductance synthesis technique is entirely based on some intuitive analysis theorems.

A Proof of the Topological Rules of Signal-Flow-Graph Analysis. Monograph No. 465 E.

A. NATHAN, E.E., M.S., D.Sc.(Eng.).

The topological rules for the calculation of a signal-flow-graph transmittance are usually proved by quite a lengthy *ad hoc* argument. This is neither necessary nor advantageous because a straightforward matrix analysis of the problem leads directly to the desired result. Such an analysis is given in the paper.

A Note on the Analysis of the Fields of Line Currents and Charges. Monograph No. 466.

P. J. LAWRENSON, M.Sc.

The method of images is widely used in the analysis of fields due to currents and charges, but it is of limited application, and a more general method is required. It is the object of the paper to show that, for the analysis of 2-dimensional fields due to linear filaments of current or charge, conformal-transformation methods can be used simply for a very wide range of problems; the only requirement is the determination of an equation (or a series of equations) which transforms a straight line into the shape of the boundary of the field. The

boundary types which can be treated simply are equipotential, flux-line, and combinations of the two. The necessary basic field solutions for each of these three types, for fields interior and exterior to given boundaries, are developed.

Use of the method is demonstrated with two examples involving simple polygonal boundaries. The first is a determination of the error caused by treating a finite boundary as if it were infinite (an approximation frequently necessary when the image method is used). The second is an estimation of the inductance of a conductor carrying high-frequency current and influenced by a conducting surface. A method of estimating the inductance of a current influenced by permeable bodies is also mentioned.

The Use of Tensor Densities in Equivalent Circuits for Field Problems. Monograph No. 467 E.

J. W. LYNN, M.Sc., Ph.D.

Equivalent electrical networks to represent fields described by certain partial differential equations have for many years been used as a standard analogue technique. In these networks the dependen and independent variables are usually represented by voltage dro and currents in the branches, called 'across' and 'through' variables and these analogue quantities on the network give the distribution o the field.

In setting up equivalent networks for this purpose Kron adde two steps to the existing concepts, namely (i) he derived circuits i terms of general co-ordinates, the topology of each network bein independent of the co-ordinate system chosen, and (ii) by applyin Stokes's theorem to the system of vectors he showed that the network thus derived could be interpreted as containing the field quantitie flowing in the filled space. The quantities measured on the networ then correspond to line, surface or volume integrals of the vecto field quantities. With this interpretation the networks become mor realistic models of the fields since the whole space is filled and th choice of co-ordinate system does not affect the physics of th phenomena.

The present paper describes the mathematical basis of this fulle interpretation of analogue networks.

PROCEEDINGS OF THE INSTITUTION OF ELECTRICAL ENGINEERS

Part B. ELECTRONIC AND COMMUNICATION ENGINEERING (INCLUDING RADIO ENGINEERING), SEPTEMBER 1961

CONTENTS

Papers for the Proceedings.—An author who supplies an outline of a paper he proposes to submit for the *Proceedings* may apply to the Secretary for a free copy of The Institution's Handbook for Authors. This gives particulars of a number of requirements—including maximum acceptable length—compliance with which is essential. See page ad 30 in the advertisement section.

Declaration on Fair Copying.—Within the terms of the Royal Society's Declaration on Fair Copying, to which The Institution subscribes, material may be copied from issues of the *Proceedings* (prior to 1949, the *Journal*) *which are out of print and from which reprints are not available.* The terms of the Declaration and particulars of a Photoprint Service afforded by the Science Museum Library, London, are published in the *Journal* from time to time.

Bibliographical References.—It is requested that bibliographical reference to an Institution paper should always include the serial number of the paper and the month and year of publication, which will be found at the top right-hand corner of the first page of the paper. This information should precede the reference to the Volume and Part.

Example.—SMITH, J.: 'Reflections from the Ionosphere', *Proceedings* I.E.E., Paper No. 5001 R, December, 1960 (102 B, p. 1234).

Published by The Institution, Savoy Place, London, W.C.2. *Telephone* COVent Garden 1871. *Telegrams:* 'Voltampere, Phone, London.
Printed by Unwin Brothers Limited, Woking and London.